

MÉTODOS BAYESIANOS EM VISÃO COMPUTACIONAL

Edison Americo Huarsaya Tito

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS EM ENGENHARIA ELÉTRICA

Aprovada por:

Prof. Antonio Petraglia, Ph.D.

Profª. Mariane Rembold Petraglia, Ph.D.

Prof. Alexandre Pinto Alves da Silva, Ph.D.

Prof. Marcelo Martins Werneck, PhD.

Prof. Basílio de Bragança Pereira, PhD.

Prof. Altamiro Amadeu Susin, PhD.

RIO DE JANEIRO, RJ - BRASIL

JUNHO DE 2005

TITO, EDISON AMERICO HUARSAYA

Métodos Bayesianos em Visão
Computacional [Rio de Janeiro] 2005

IX, 66 p. 29,7 cm (COPPE/UFRJ,
D.Sc., Engenharia Elétrica, 2005)

Tese - Universidade Federal do Rio de
Janeiro, COPPE

1. Estatística Bayesiana 2. Visão
Computacional

I. COPPE/UFRJ II. Título (série)

A meus pais e meus irmãos

Agradecimentos

Aos meus orientadores, Dr. Antonio Petraglia e Dra. Mariane Rembold de Petraglia pelo apoio, amizade, confiança depositada e incentivo, com que pude contar em todos os momentos da realização desse trabalho.

Ao CNPq e à UFRJ, pelos auxílios concedidos, sem os quais este trabalho não poderia ter sido realizado.

Aos professores, pesquisadores, funcionários e colegas do Departamento de Engenharia Elétrica da UFRJ.

Aos professores que participaram da comissão examinadora.

Aos meus pais, pela educação, atenção e carinho de todas as horas.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

MÉTODOS BAYESIANOS EM VISÃO COMPUTACIONAL

Edison Americo Huarsaya Tito

Junho/2005

Orientadores: Antonio Petraglia
Mariane Rembold Petraglia

Programa: Engenharia Elétrica

Este trabalho tem por objetivo estudar o problema de visão artificial não simplesmente imitando as capacidades da visão humana e sim imitando a natureza tanto no problema quanto na solução.

Imitar a natureza no problema implica em tratar os sinais visuais provenientes do mundo exterior como eles realmente são: sinais não-lineares, não gaussianos, não estacionários e variantes no tempo.

Imitar a natureza na solução implica em evoluir o modelo de visão computacional, selecionando e ajustando o melhor modelo, sob a ótica da genética Mendeliana e do princípio Darwiniano da “sobrevivência dos mais aptos”.

Como contribuição principal deste trabalho destaca-se o estabelecimento de uma equivalência entre o princípio Darwiniano e o teorema de Bayes, permitindo adaptar os métodos convencionais de simulações estocásticas de amostragem ou de otimização numa visão Mendeliana - Darwiniana.

Como resultado obtêm-se métodos de simulações estocásticas híbridos capazes de ajustar modelos probabilísticos complexos adequados para o processamento de sinais do mundo real.

Para a comprovação das idéias expostas neste trabalho, foram desenvolvidas duas aplicações de visão artificial com resultados satisfatórios, mostrando que a metodologia é viável de ser implementada na prática em aplicações reais.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

BAYESIAN METHODS FOR COMPUTER VISION

Edison Americo Huarsaya Tito

June/2005

Advisors: Antonio Petraglia
Mariane Rembold Petraglia

Department: Electrical Engineering

The aim of this work is to study the problem of artificial vision not simply imitating the capabilities of the human vision, but mainly imitating nature in the problem as well as in the solution.

Imitate nature in the problem implies in treating the visual signals proceeding from the external world as they really are: nonlinear, non-Gaussian, non-stationary and time-variant.

Imitate nature of the solution implies in evolving the model of artificial vision, selecting and adjusting the best model, front the standpoint of the Mendelian genetics and the Darwinian principle of the "survival of the fittest".

The main contribution of this work is the establishment of the equivalence between the Darwinian principle and the Bayes' theorem, allowing to adapt the conventional methods of stochastic simulation of sampling or optimization into a Mendelian - Darwinian vision.

As a result, we obtain hybrid stochastic simulation methods capable of adjusting probabilistic complex models, which are adequate for processing real world signals.

To verify the ideas proposed in this work, two applications of artificial vision with satisfactory results have been developed, showing that the developed methodology is viable in real applications.

Sumário

1	Introdução	1
1.1.	Motivação	1
1.2.	Objetivos e Contribuições deste Trabalho	4
1.3.	Organização da Tese	5
2	A Natureza da Visão Humana	6
2.1.	Teoria da Seleção Natural dos Replicadores	7
2.2.	Teoria Computacional da Mente Humana	9
2.3.	Extensão do Princípio Darwiniano em Ambientes probabilísticos	10
2.4.	Resumo	25
3	Visão Artificial por Métodos Computacionais	27
3.1.	Modelo Probabilístico de Visão Artificial	28
3.2.	Otimização Global do Sistema de Visão Artificial	40
4	Aplicações do Modelo de Visão Artificial	44
4.1.	Aplicação de Visão de Objetos Tridimensionais	44
4.2.	Aplicação de Visão Submarina na Indústria de Petróleo	46
4.3.	Detalhes da Implementação Computacional do Algoritmo	57
5	Conclusões e Trabalhos Futuros	61
6	Referências Bibliográficas	63

Lista de figuras

Figura 1. Problema de reconhecimento/classificação de padrões.	2
Figura 2. Sistema de visão computacional.	3
Figura 3. Função de custo a ser minimizada.	11
Figura 4. Notação simbólica da evolução de cromossomas.	11
Figura 5. Função de custo a ser maximizada.	12
Figura 6. Evolução da distribuição marginal $p(\mathbf{x}_t)$.	19
Figura 7. Evolução da distribuição marginal $p(\mathbf{y}_t)$.	20
Figura 8. Convergência da moda da distribuição $p(\boldsymbol{\theta}_t) = p(\mathbf{x}_t, \mathbf{y}_t)$.	20
Figura 9. Equivalência entre o princípio Darwiniano e o teorema de Bayes.	24
Figura 10. Paradigmas de amostragem e/ou otimização.	25
Figura 11. Problema da Profundidade.	27
Figura 12. Ilusão óptica: Triângulo de Kanisza.	28
Figura 13. Reconhecimento de objetos tridimensionais em ambientes dinâmicos	28
Figura 14. Sistema mono ocular com formato de cores em tons de cinza.	29
Figura 15. Representação vetorial de uma imagem.	30
Figura 16. Curva formada pelas infinitas vistas do objeto.	30
Figura 17. Modelo de visão computacional proposto neste trabalho.	31
Figura 18. Fluxo dos sinais através dos módulos do sistema.	32
Figura 19. Modelo paramétrico de sistema de visão artificial.	32
Figura 20. Parâmetros do Sistema Mono Ocular.	33
Figura 21. Concentração e dispersão das imagens.	34
Figura 22. Redução de dimensionalidade (eliminação do eixo x_2).	35
Figura 23. Parâmetros do módulo de redução de dimensionalidade.	36
Figura 24. Arquitetura Perceptron com Duas Camadas.	37
Figura 25. Função de ativação Logística Sigmoidal.	38
Figura 26. Parâmetros e Hiper-parâmetros do módulo de RNA.	40
Figura 27. Aplicativo de renderização de um cone tridimensional.	44
Figura 28. Vistas do cone tridimensional.	45
Figura 29. Recordes da Petrobras na exploração <i>offshore</i> .	46
Figura 30. Equipamentos num campo de produção de petróleo.	47

Figura 31. Equipamento de cabeça de poço.	48
Figura 32. Imagem filtrada.	48
Figura 33. Detecção de borda dos objetos.	49
Figura 34. Isolamento do objeto de interesse.	49
Figura 35. Padrões da primeira vista do objeto.	50
Figura 36. Padrões da segunda vista do objeto.	51
Figura 37. Padrões da terceira vista do objeto.	52
Figura 38. Padrões da quarta vista do objeto.	53
Figura 39. Padrões da quinta vista do objeto.	54
Figura 40. Reconhecimento on-line do objeto.	56
Figura 41. Detalhes dos Parâmetros do Sistema de Visão Artificial.	58
Figura 42. Otimização do Sistema de Sistema de Visão Artificial.	58

1 Introdução

Qualquer trabalho que trate de reproduzir de forma artificial alguma das faculdades do ser humano, como é o caso da faculdade da visão, deveria começar com uma nota de humildade; este trabalho começara com duas.

Primeiro, a ciência não compreende totalmente o funcionamento do corpo e a mente do ser humano; uma máquina de sobrevivência incrivelmente adaptativa, versátil, complexa e misteriosa. Porém, este trabalho de visão artificial baseado em métodos computacionais – visão computacional – deve ser considerada apenas uma incompleta e modesta abstração computacional da visão humana.

Em segundo lugar, este trabalho não pretende explicar o funcionamento da visão humana em termos de algoritmos computacionais, apenas procura deduzir estes algoritmos de maneira coerente com base nos conhecimentos acumulados pelas diversas disciplinas e teorias da ciência relacionadas com o tema.

O restante deste capítulo inicial está dividida em três seções. A primeira seção descreve as motivações que deram lugar à esta pesquisa. A segunda seção enumera os objetivos e contribuições deste trabalho e na última seção, descreve-se a organização do restante da tese.

1.1. Motivação

O advento de uma formidável capacidade computacional de processamento de dados a preços acessíveis tem estimulado o interesse da indústria pelo desenvolvimento de sistemas automáticos que visem, entre outros, substituir o ser humano em tarefas consideradas perigosas, mecânicas e/ou repetitivas. Como exemplos destes sistemas pode-se citar: sistemas de visão submarina, sistemas de vigilância, sistemas de reconhecimento de fase, impressão digital, identificação de locutor, sistemas de controle de processos manufaturados, sistemas de reconhecimento e seguimento de objetos 3D, etc. Estes sistemas, em essência,

procuram imitar de forma artificial certas habilidades humanas tais como: visão, audição, paladar, olfato, tato, coordenação motora, bom senso, raciocínio, etc.

Embora a concepção do problema para estes sistemas esteja claramente definida, a procura pela sua solução resultou ser um enorme desafio. Diversas áreas da ciência já tem contribuído com estudos focalizados sobre este problema, entretanto, a área de *Inteligência Artificial* é a área que tem abraçado a missão da procura da solução deste problema, para isto, conta inclusive com linhas auxiliares de pesquisa destas das quais destaca-se a linha de *reconhecimento de padrões* isto porque um percentual elevado destes sistemas tem como característica principal a de reconhecer um determinado padrão de entrada, como por exemplo, padrões de voz, padrões de impressão digital, padrões de fases, padrões de objetos, etc. Apesar do fato de que esta linha de pesquisa tenha surgido há quase 50 anos, o problema genérico de reconhecimento de padrões com orientação, posição e escala arbitrária, continua até hoje sem solução [1].

O termo *padrão* é definido na comunidade de reconhecimento de padrões “como o oposto ao caos, uma entidade vagamente definida à qual pode ser atribuído um nome” [1]. Além disso, esta comunidade formula o problema de *reconhecimento de padrões* como um problema de *classificação de padrões* [2] (Figura 1), onde as classes são previamente definidas pelo projetista do sistema (classificação supervisionada) ou são aprendidas/extraídas baseadas na similaridade dos padrões (classificação não supervisionada).

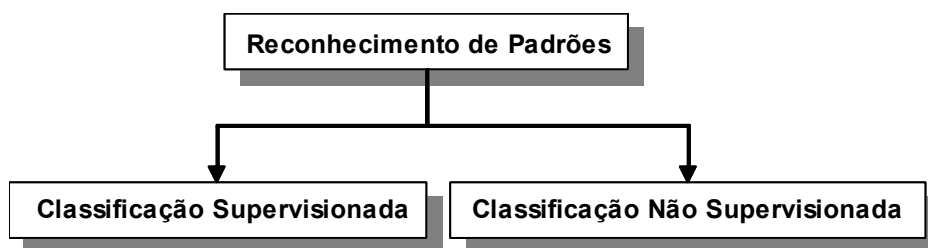


Figura 1. Problema de reconhecimento/classificação de padrões.

Neste contexto, o projeto da maioria dos sistemas de reconhecimento de padrões envolve essencialmente as seguintes três etapas: (a) aquisição de dados; (b) pré-processamento de dados e (c) estratégia de classificação.

Para o caso dos sistemas de visão computacional (Figura 2), a etapa de aquisição de dados, refere-se ao processo de captura e codificação dos sinais do

mundo real numa seqüência temporal de imagens digitalizadas; a etapa de pré-processamento de dados refere-se ao processo de padronização das imagens e à extração de características (métricas) dos objetos contidos nestas imagens padronizadas e a terceira etapa compreende a formulação do modelo de classificação a ser empregado como suporte a decisão.



Figura 2. Sistema de visão computacional.

Diversas metodologias de classificação foram propostas tais como: janelas de Parzen, Discriminante de Fisher, classificador Fuzzy, redes neurais, etc. (detalhes destes métodos em [2]-[6]). Analogamente diversas metodologias de pré-processamento foram propostas tais como: análise de componentes independentes [7], análise de componentes principais [8], redes neurais, etc. Além de toda a teoria de processamento de sinais [9][10] destinadas ao processamento e filtragem das imagens.

Em resumo, existem diversas metodologias procurando otimizar o desempenho de cada módulo ou sub-módulo do sistema de forma isolada. Entretanto, existem poucas metodologias proposta que visem um estudo do sistema de forma global e integrada, o qual é fundamental, pois a otimização do desempenho de cada módulo e sub-módulo isoladamente não garante a otimização do desempenho de todo o sistema.

Além disso, as metodologias propostas para cada módulo têm forte inspiração em conceitos matemáticos abstratos o qual representa uma aparente contradição já que a inspiração principal deveria ser imitar natureza uma vez que nosso objetivo é imitar as habilidades da visão humana.

Neste contexto, a motivação do presente trabalho é abordar o problema de visão artificial, não simplesmente imitando as habilidades da visão humana, mas também imitando seu processo evolutivos de construção, procurando desta forma uma solução integrada e otimizada globalmente.

Conseqüentemente, o foco principal deste trabalho recai na procura de uma metodologia global de otimização do sistema de visão computacional. Cabe destacar que os modelos matemáticos de cada sub-módulo do sistema foram

selecionados com base a modelos consagrados na literatura ou selecionados por ser modelos inspirados na natureza.

Com estas considerações a seguir são enumerados os objetivos e contribuições deste trabalho.

1.2. Objetivos e Contribuições deste Trabalho

A seguir são enumerados, de forma genérica, os objetivos e contribuições deste trabalho. Entretanto, no último capítulo voltaremos a enumerar as contribuições da tese, mas de uma forma detalhada e precisa com base nos resultados obtidos e nos argumentos desenvolvidos ao longo do trabalho.

1.2.1. Objetivos

- Abordar o problema de visão artificial não simplesmente imitando as habilidades da visão humana e sim imitando a natureza tanto no problema quanto na solução.
- Imitar a natureza no problema implica em formular o modelo de visão computacional com a capacidade de processar sinais visíveis complexos do mundo real, caracterizados por serem ruidosos, dinâmicos e variantes no tempo.
- Imitar a natureza na solução implica em evoluir o modelo de visão computacional, selecionando e ajustando o melhor modelo, sob a ótica da genética Mendeliana e do princípio Darwiniano da “sobrevivência dos mais aptos”.
- Comprovar através de estudo de casos a eficiência e robustez do modelo de visão computacional proposto no presente trabalho.

1.2.2. Contribuições

- A formulação de um modelo de visão computacional inspirado na natureza tanto no problema quanto na solução, abordando desta forma o problema de visão computacional de uma forma global, integrada e coerente com os conhecimentos da ciência sobre o tema.

- A formulação de um modelo probabilístico de visão computacional capaz de processar sinais visíveis (sinais de vídeo) complexos do mundo real caracterizados por comportamentos não-lineares, não-gaussianos, não-estacionários e dinâmico no tempo.
- O estabelecimento de uma equivalência entre o princípio Darwiniano (unido com a genética Mendeliana) e o teorema de Bayes, permitindo adotar a estatística Bayesiana como plataforma de análise e síntese do modelo probabilístico de visão computacional. Além disso, esta equivalência permite adaptar determinados métodos de simulações estocásticas de amostragem (aproximação numérica da densidade *a posteriori*) ou de otimização (aproximação da moda da distribuição *a posteriori*) numa visão Mendeliana/ Darwiniana.
- Com a comprovação do modelo de visão computacional proposto, foi desenvolvida com sucesso uma aplicação real em visão submarina.

1.3.

Organização da Tese

O restante desta tese está organizada em mais quatro capítulos, discutidos a seguir:

No Capítulo 2 apresenta-se brevemente a natureza da visão humana, assim como uma modelagem matemática da teoria de evolução de Darwin..

No Capítulo 3 apresenta-se o modelo de visão computacional proposto.

No Capítulo 4 o modelo de visão computacional aqui proposto é utilizado em duas aplicações: (a) no reconhecimento de objetos tridimensionais e (b) visão submarina na indústria de petróleo. Finalmente, o Capítulo 5 apresenta-se às conclusões e propostas de trabalhos futuros.

2

A Natureza da Visão Humana

Antes de deduzir as premissas do nosso modelo de visão computacional, é importante lembrar que nosso foco principal é a “otimização global” do modelo de visão computacional. Conseqüentemente, nosso problema pode ser visto como um problema de estimação de modelos (nomenclatura estatística), aprendizado de modelos (nomenclatura de visão artificial) ou filtragem de ruído (nomenclatura de processamento de sinais). Independentemente da nomenclatura, nesta abordagem do problema pressupõe-se que temos um modelo matemático parametrizável a qual precisa ser ajustado a uma serie histórica de dados (nomenclatura de estatística), conjunto de treinamento (nomenclatura de inteligência artificial) ou um sinal (nomenclatura de processamento de sinais).

Em contrapartida existe uma outra abordagem do problema a qual pretende reproduzir matematicamente cada componente que realiza a função da visão humana. Nesta linha de pensamento destaca-se David Marr [11], que demonstrou por exemplo, que os contornos de um objeto podem ser detectados como a raiz $\nabla^2 \mathbf{G}(x, y) = 0$ do Laplaciano $\nabla^2 (\partial^2 / \partial x^2 + \partial^2 / \partial y^2)$ da função da intensidade de iluminação $\mathbf{G}(x, y)$ da imagem definida pelos pixels (x, y) . Também mostrou que a geometria do campo receptivo da retina pode ser construída como a diferença de duas Gaussianas, etc. Pesquisas nesta linha encontram-se em [12],[13].

Com base nestas considerações a abordagem deste trabalho focaliza-se na estimação de modelos. Para isto é necessário estudar a natureza da visão humana do ponto de vista da teoria da seleção natural dos replicadores [14], e a teoria computacional da mente [15]. Estas duas teorias se relacionam com o foco deste trabalho que é a “otimização global do sistema de visão artificial”, ou seja, o ajuste do modelo de visão computacional.

Começaremos com a premissa de que somos os únicos seres inteligentes no planeta Terra, e como tal tentamos procurar sempre respostas racionais a qualquer pergunta, como por exemplo: Qual é a origem homem? Para esta pergunta, em particular, a teoria da evolução por *seleção natural* de Charles Darwin [16]

fornece uma resposta racional e coerente, uma vez que mostra como a simplicidade poder-se-ia transformar em complexidade, como átomos desordenados poderiam se agrupar em padrões cada vez mais complexos, até que terminassem por fabricar pessoas.

Atualmente, esta teoria está sujeita a dúvidas e/ou completamente ignoradas, mas as implicações plenas da revolução de Darwin estão por ser amplamente compreendidas. Em essência, o pensamento de Darwin defende que a evolução trabalha através da “seleção natural” e esta significa a “sobrevivência dos mais aptos”. Mas estamos falando sobre espécies mais aptas, raças mais aptas, indivíduos mais aptos, ou sobre o quê? Este é um ponto crucial para os biólogos, pois travam verdadeiras batalhas, defendendo a “seleção de grupo” ou a “seleção individual” como unidade fundamental de seleção. Entretanto, as evidências mostram que a evolução é uma propriedade do cosmos pela qual todas as entidades que conformam seus diversos ambientes estão em constante mudança na procura de um ponto de equilíbrio, na procura pela estabilidade¹. Esta é a revolução introduzida por Darwin que precisa ser entendida. Desta forma, a seleção de grupo ou a seleção individual podem perfeitamente ser consideradas unidades de seleção. Entretanto, quando pretendemos entender nossa origem, entender como obtivemos nosso corpo e porque temos esta mente específica com todas suas habilidades, a unidade fundamental de seleção que fornece as respostas mais convincentes é a seleção individual, mais especificamente a seleção de *genes*. Esta variante da teoria de evolução é conhecida como a teoria da seleção natural dos replicadores a qual será abordada brevemente na próxima seção.

2.1. Teoria da Seleção Natural dos Replicadores

Calcula-se que quatro bilhões de anos atrás surgiram no mar, pela primeira vez, moléculas com a faculdade de fazer cópias de si próprias. Tão logo o *replicador* molecular surgiu, ele deve ter espalhado suas cópias rapidamente pelos mares. À medida que cópias errôneas eram introduzidas e propagadas pelo processo de replicação, o mar encheu-se de uma população, não de réplicas idênticas, mas de variedades de moléculas replicadoras, todas descendentes do

¹ Analogamente ao princípio de entropia.

mesmo ancestral. À medida que os blocos de construção destas moléculas replicadoras passaram a ser um recurso escasso e precioso, houve uma luta pela sobrevivência entre as distintas variedades de replicadores. É importante destacar que estes replicadores não sabiam que estavam lutando, nem se preocupavam com isto. A luta foi conduzida sem quaisquer maus sentimentos, de fato sem sentimentos de qualquer espécie. Algumas variedades talvez até tenham descoberto como quebrar quimicamente as moléculas de linhagens rivais e utilizar os constituintes assim liberados para fazer suas próprias cópias. Outros replicadores talvez tenham descoberto como se proteger, quer quimicamente, quer construindo uma parede física de proteína ao redor de si. Talvez tenha sido assim que as primeiras células vivas apareceram. Os replicadores começam não apenas a existir, mas a construir envoltórios para si, veículos para sua existência ininterrupta. Assim, os replicadores bem mais sucedidos foram aqueles que desenvolveram as *máquinas de sobrevivência* mais sofisticadas. Quatro bilhões de anos mais tarde, qual seria o destino destes antigos replicadores? Que máquinas estranhas de auto-sobrevivência todos esses milênios produziram? A resposta é que esses replicadores não foram extintos e agora eles são conhecidos pelo nome de *genes* e nós somos suas *máquinas de sobrevivência*.

É desta maneira que Richard Dawkins [14] interpreta o princípio Darwiniano da “sobrevivência dos mais aptos” e nos mostra como nosso corpo alcançou extrema perfeição e complexidade. Neste contexto, fica claro que as diversas faculdades do corpo humano, como por exemplo a visão humana, foram projetadas para colaborar com a nossa máquina de sobrevivência na preservação e na reprodução de nossos genes.

Apesar de que a teoria da seleção natural dos replicadores explique convincentemente as nossas origens, ela não explica como é a conexão entre o etéreo mundo dos significados, a essência da nossa vida mental, a um pedaço físico de matéria como é o cérebro. Esta conexão é explicada pela teoria computacional da mente, assunto este que será abordado brevemente na seção seguinte.

2.2. Teoria Computacional da Mente Humana

A teoria computacional da mente, afirma que as diversas entidades do mundo exterior desencadeiam informações como configurações de símbolos. Estes símbolos são mapeados via órgãos dos sentidos ao cérebro como *padrões de dados*. A mente processa estes padrões de dados contrastando-os com outros padrões de dados previamente armazenados. Como resultado destas operações, pode ser desencadeado no corpo um comportamento, ou seja uma determinada ação física. Além disso, a teoria computacional da mente defende também que a mente humana é um sistema de órgãos de computação, projetados pela seleção natural para permitir aos nossos ancestrais o domínio sobre pedras, utensílios, plantas e animais, em última análise a serviço da sobrevivência e reprodução. A mente está organizada em módulos ou órgãos mentais, cada qual com um desenho especializado que faz desse módulo um perito em uma área de interação com o mundo. A lógica básica desses módulos é especificada no nosso programa genético. Para esta teoria a mente não é o cérebro e sim o que o cérebro faz. Para um melhor entendimento pode-se estabelecer uma analogia direta entre cérebro/mente com hardware/software de um computador.

Para o caso específico da visão humana, é fundamental considerar que estes símbolos ou padrões de dados estão sujeitos geralmente a perturbações ruidosas. Isto porque segundo a teoria quântica [17], o universo está governado, entre outros, pelo princípio de incerteza de Heisenberg, que demonstrou que a incerteza está inserida na natureza desde seus níveis atômicos mais fundamentais. Este princípio teve profundas implicações na forma de percepção do mundo, e assinalou o fim do sonho de uma teoria da ciência para um modelo do universo completamente determinístico; isto porque não é possível prever eventos futuros com precisão uma vez que não sendo possível medir precisamente o estado presente do universo, sempre teremos uma incerteza ou um ruído remanescente.

Desta forma além teoria da seleção natural dos replicadores e a teoria computacional da mente é necessário adicionar a teoria de probabilidades [18] como a plataforma de análise do problema de visão artificial. Para isto, na seção seguinte, propõe-se uma extensão do princípio Darwiniano em ambientes probabilísticos.

2.3.

Extensão do Princípio Darwiniano em Ambientes Probabilísticos

A teoria da seleção natural dos replicadores, exposta na primeira seção, apresentou uma variante da teoria de Darwin que tenta explicar de forma mais precisa toda a seqüência da evolução dos seres vivos no planeta, partindo de estruturas atômicas simples até organismos tão complexos como o ser humano. Entretanto, as evidências mostram que as idéias evolutivas de Darwin transcendem além do mundo orgânico, para ambientes inorgânicos e de maneira geral para qualquer ambiente do cosmos onde as entidades que o conformam também evoluem e isto significa no seu sentido mais genérico, tentativas de alcançar estruturas mais estáveis ou configurações de equilíbrio. Generalizando este raciocínio, John Holland aplicou com sucesso as idéias de Darwin em ambientes abstratos como é o caso de problemas de otimização. Seu trabalho, publicado em 1975, intitulado “Adaptation in Natural and Artificial Systems” [19], é considerado a semente de uma nova tecnologia de otimização denominada de Algoritmos Genéticos (AG) [20][21].

A extensão do princípio Darwiniano em ambientes probabilísticos pode ser deduzida a partir da aplicação do princípio Darwiniano em problemas de otimização, desenvolvido por Holland e exposta brevemente a seguir.

Seja $E(\boldsymbol{\theta}) \geq 0, \forall \boldsymbol{\theta} \in \mathfrak{R}^n$ uma determinada função de custo a ser minimizada. Então, o problema de otimização consiste em alcançar o ponto $\boldsymbol{\theta}^{\min}$, definido pela equação:

$$\boldsymbol{\theta}^{\min} = \arg \min_{\boldsymbol{\theta}} [E(\boldsymbol{\theta})] \quad (1)$$

Na qual a função $E(\boldsymbol{\theta})$ adota seu valor mínimo global. Como exemplo ilustrativo, a figura (3) mostra a superfície da função de custo a ser minimizada definida por $E(\boldsymbol{\theta}) = E(\mathbf{x}, \mathbf{y}) = 0.01(\mathbf{x}^2 + \mathbf{y}^2)$. Para esta função de custo seu valor mínimo global ocorre em $\boldsymbol{\theta}^{\min} = [\mathbf{x}, \mathbf{y}] = [0, 0]$.

Holland abordou esta classe de problemas, primeiramente, visualizando o vetor $\boldsymbol{\theta} \in \mathfrak{R}^n$ de n-componentes dimensionais, como se fosse um cromossoma de n-genes. Para o exemplo anterior, qualquer par de coordenadas $\boldsymbol{\theta}^{(i)} = [\mathbf{x}^{(i)}, \mathbf{y}^{(i)}]$ seria um cromossoma e cada coordenada representaria um gene.

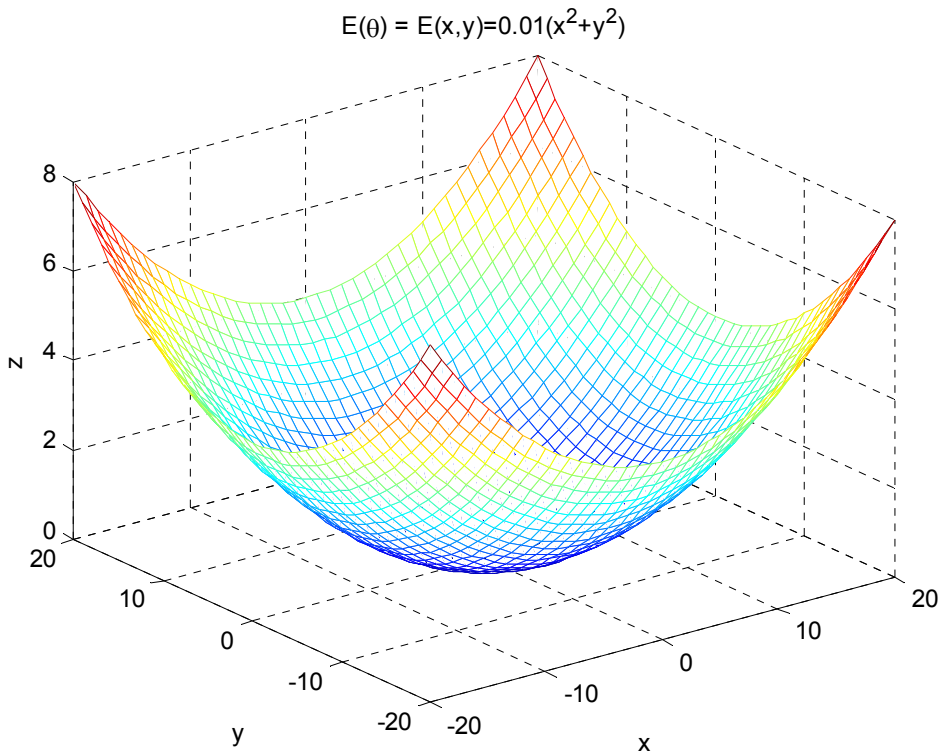


Figura 3. Função de custo a ser minimizada.

Holland, na seqüência, visualizou a função $E(\theta)$ como se fosse um ambiente Darwiniano artificial, na qual seria possível evoluir uma população de N cromossomas $\{\theta_{0t}^{(i)}\}_{i=1}^N = \{\theta_0^{(i)}\}_{i=1}^N, \{\theta_1^{(i)}\}_{i=1}^N, \dots, \{\theta_k^{(i)}\}_{i=1}^N, \dots, \{\theta_t^{(i)}\}_{i=1}^N$ ao longo de t gerações. Neste contexto, Holland previu que o indivíduo mais apto deste ambiente artificial, θ^{\min} , poderia ser encontrado seguindo os princípios da genética Mendeliana e do princípio Darwiniano da “sobrevivência dos mais aptos”.

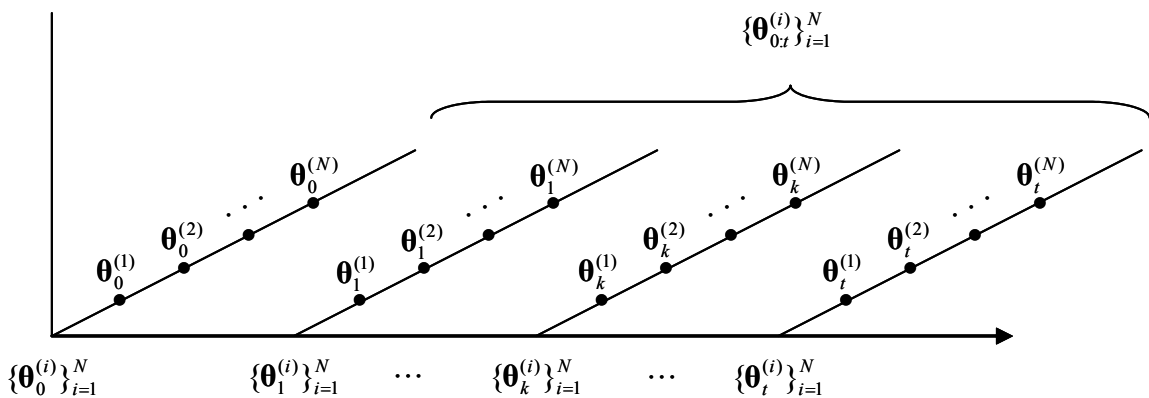


Figura 4. Notação simbólica da evolução de cromossomas.

A notação $\{\boldsymbol{\theta}_{0,t}^{(i)}\}_{i=1}^N$ representa todos os cromossomas, desde a população da geração inicial $\{\boldsymbol{\theta}_0^{(i)}\}_{i=1}^N$ até a população da geração atual $\{\boldsymbol{\theta}_t^{(i)}\}_{i=1}^N$. A população da k -ésima geração, como qualquer outra geração, está composta de N cromossomas $\{\boldsymbol{\theta}_k^{(i)}\}_{i=1}^N = \{\boldsymbol{\theta}_k^{(1)}, \boldsymbol{\theta}_k^{(2)}, \dots, \boldsymbol{\theta}_k^{(N)}\}$. A figura (4) mostra graficamente esta notação.

Sem perda de generalidade, qualquer problema de minimização pode ser reformulado como um problema de maximização. Para os propósitos deste trabalho a função de custo $E(\boldsymbol{\theta})$ pode ser redefinida convenientemente através da seguinte transformação monotônica:

$$p(\boldsymbol{\theta}) = Z^{-1} \exp[-E(\boldsymbol{\theta})] \quad (2)$$

onde $Z^{-1} = \int \exp[-E(\boldsymbol{\theta})] d\boldsymbol{\theta}$. Para esta nova função de custo $p(\boldsymbol{\theta})$, o problema de otimização, transforma-se na procura do ponto $\boldsymbol{\theta}^{\max}$, definido por:

$$\boldsymbol{\theta}^{\max} = \arg \max_{\boldsymbol{\theta}} [p(\boldsymbol{\theta})] \quad (3)$$

Na qual a função $p(\boldsymbol{\theta})$ assume seu valor máximo global. A transformação monotônica da equação (2) garante que as soluções definidas pelas equações (1) e (3) sejam idênticas, $\boldsymbol{\theta}^{\min} \equiv \boldsymbol{\theta}^{\max}$. Para o exemplo de otimização acima descrito, a figura (5) mostra a superfície da função de custo equivalente a ser maximizada.

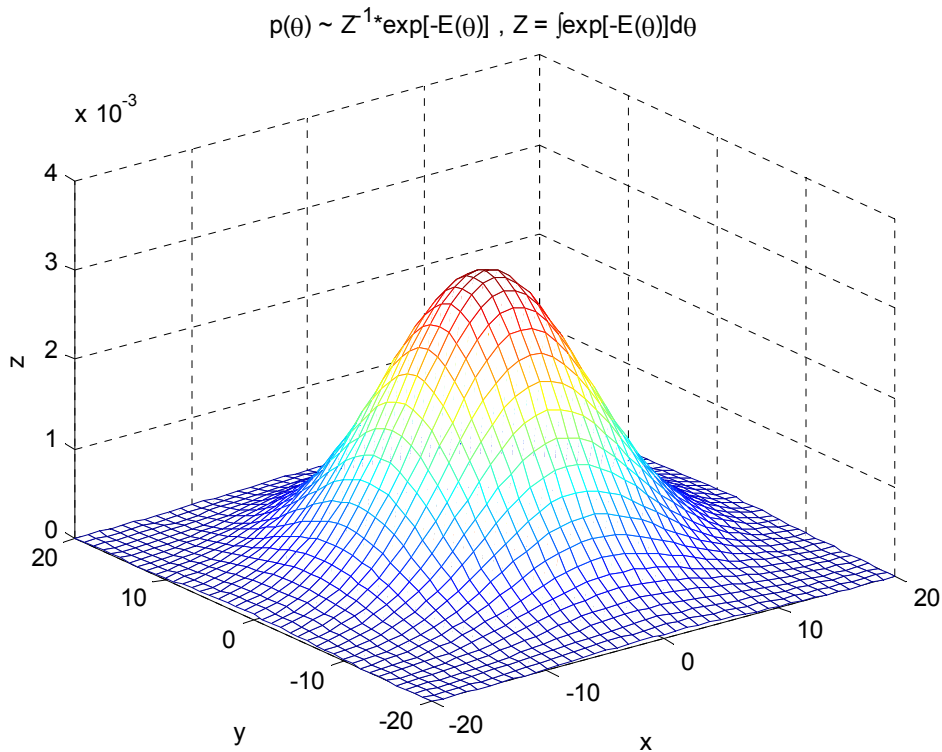


Figura 5. Função de custo a ser maximizada.

O relevante desta transformação é que a nova função de custo $p(\boldsymbol{\theta})$ apresenta, notavelmente, todas as propriedades de qualquer função de probabilidade. Portanto, o processo evolutivo da população de cromossomas $\{\boldsymbol{\theta}_{0:t}^{(i)}\}_{i=1}^N$ equivaleria, em termos de simulações estocásticas, a um processo de “amostragem” da função de distribuição $p(\boldsymbol{\theta})$ com a intenção de “procurar” sua moda $\boldsymbol{\theta}^{\max}$ global.

Este tipo de amostradores/otimizadores estocásticos é de extrema importância, pois a solução de inúmeros problemas da engenharia, como é caso do problema de visão artificial, não admite um tratamento analítico fechado. Nestes casos, muitas vezes, a solução reduz-se a um problema de otimização ou de amostragem de funções de probabilidade.

Devido à esta importância, a seguir derivam-se as condições suficientes que garantem a convergência dos algoritmos computacionais baseados nesta nova extensão do princípio Darwiniano. Também deriva-se o pseudocódigo do algoritmo a ser implementado computacionalmente.

2.3.1. Condições de Convergência

O processo evolutivo da população de cromossomas pode ser modelado probabilisticamente através de um processo “dinâmico” Markoviano $\pi(\boldsymbol{\theta}_{0:t})$ definido num espaço dimensional de variáveis aleatórias $\boldsymbol{\theta}_{0:t} = \{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_t\}$ crescente com o tempo. Analogamente, a função de distribuição “objetivo” $p(\boldsymbol{\theta})$, pode ser usado para definir um processo estocástico “estacionário” $p(\boldsymbol{\theta}_{0:t})$ imerso no mesmo espaço dimensional $\boldsymbol{\theta}_{0:t} = \{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_t\}$ de variáveis aleatórias, mas com a propriedade de que suas densidades marginais $p(\boldsymbol{\theta}_0), p(\boldsymbol{\theta}_1), \dots, p(\boldsymbol{\theta}_t)$ sejam “invariantes no tempo” e todas iguais a $p(\boldsymbol{\theta})$.

Neste contexto, o processo evolutivo Darwiniano $\pi(\boldsymbol{\theta}_{0:t})$ tem como missão aproximar-se suavemente do processo objetivo $p(\boldsymbol{\theta}_{0:t})$, conforme o transcurso do tempo t , e com isto estimar algumas quantidades desconhecidas de $p(\boldsymbol{\theta}_{0:t})$, tal

como sua moda global $\boldsymbol{\theta}^{\max}$. Uma forma de monitorar a convergência de $\pi(\boldsymbol{\theta}_{0:t})$ para $p(\boldsymbol{\theta}_{0:t})$ é através da distância *Chi-square* [22], definida como a variância $\text{var}_{\pi}[\omega(\boldsymbol{\theta}_{0:t})]$ de $\omega(\boldsymbol{\theta}_{0:t})$, definida por

$$\omega(\boldsymbol{\theta}_{0:t}) = \frac{p(\boldsymbol{\theta}_{0:t})}{\pi(\boldsymbol{\theta}_{0:t})} \quad (4)$$

Tipicamente, $\omega(\boldsymbol{\theta}_{0:t})$ é denominada como o *peso de importância* a qual pode ser vista em termos Darwinianos como a *aptidão* do cromossoma. Contudo, grandes valores da distância *Chi-square* indicam uma aproximação deficiente de $\pi(\boldsymbol{\theta}_{0:t})$ com relação a $p(\boldsymbol{\theta}_{0:t})$. Na prática, a distância *Chi-square* não pode ser calculada analiticamente. Felizmente uma aproximação numérica baseada no *coeficiente de variação*, $C^2(\omega_t)$, foi sugerida em [22],

$$C^2(\tilde{\omega}) = \frac{1}{N} \sum_{i=1}^N \{N\tilde{\omega}_t^{(i)} - 1\} \quad (5)$$

onde $\tilde{\omega}_t^{(i)}$ é denominado de *peso de importância normalizado* ou de *aptidão normalizada*; e definida por

$$\tilde{\omega}_t^{(i)} = \frac{\omega(\boldsymbol{\theta}_{0:t}^{(i)})}{\sum_{j=1}^N \omega(\boldsymbol{\theta}_{0:t}^{(j)})} \quad (6)$$

Com base neste conjunto de amostras ponderadas $\{\boldsymbol{\theta}_{0:t}^{(i)}, \tilde{\omega}_t^{(i)}\}_{i=1}^N$, a densidade $p(\boldsymbol{\theta}_{0:t})$ pode ser aproximada empiricamente por

$$\hat{p}(\boldsymbol{\theta}_{0:t}) = \frac{1}{N} \sum_{i=1}^N \delta_{\boldsymbol{\theta}_{0:t}^{(i)}}(\boldsymbol{\theta}_{0:t}) \quad (7)$$

onde $\delta_{\boldsymbol{\theta}_{0:t}^{(i)}}(\boldsymbol{\theta}_{0:t})$ denota a massa do delta-Dirac localizado em $\boldsymbol{\theta}_{0:t}^{(i)}$. Uma forma eficiente de estimar $\tilde{\omega}_t^{(i)}$ ao longo do tempo é através do algoritmo de *amostragem por importância* [23]-[25], a qual aproveita a estrutura recursiva da equação (4)

$$\begin{aligned} \omega(\boldsymbol{\theta}_{0:t}) &= \frac{p(\boldsymbol{\theta}_0)p(\boldsymbol{\theta}_1|\boldsymbol{\theta}_0)\cdots p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_0,\dots,\boldsymbol{\theta}_{t-1})}{\pi(\boldsymbol{\theta}_0)\pi(\boldsymbol{\theta}_1|\boldsymbol{\theta}_0)\cdots\pi(\boldsymbol{\theta}_t|\boldsymbol{\theta}_0,\dots,\boldsymbol{\theta}_{t-1})} \\ &= \frac{p(\boldsymbol{\theta}_{0:t-1})p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_0,\dots,\boldsymbol{\theta}_{t-1})}{\pi(\boldsymbol{\theta}_{0:t-1})\pi(\boldsymbol{\theta}_t|\boldsymbol{\theta}_0,\dots,\boldsymbol{\theta}_{t-1})} \\ &= \omega(\boldsymbol{\theta}_{0:t-1}) \frac{p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_0,\dots,\boldsymbol{\theta}_{t-1})}{\pi(\boldsymbol{\theta}_t|\boldsymbol{\theta}_0,\dots,\boldsymbol{\theta}_{t-1})} \end{aligned} \quad (8)$$

Substituindo a equação acima na equação (6), o *peso de importância normalizado* satisfaz à seguinte forma recursiva:

$$\begin{aligned}\tilde{\omega}_t^{(i)} &\propto \tilde{\omega}_{t-1}^{(i)} \frac{p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_0, \dots, \boldsymbol{\theta}_{t-1})}{\pi(\boldsymbol{\theta}_t | \boldsymbol{\theta}_0, \dots, \boldsymbol{\theta}_{t-1})} \\ &\propto \tilde{\omega}_{t-1}^{(i)} \frac{p(\boldsymbol{\theta}_t)}{\pi(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1})}\end{aligned}\tag{9}$$

Na equação anterior é preciso lembrar que o processo estocástico estacionário $p(\boldsymbol{\theta}_{0:t})$ é invariante no tempo, então $p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_0, \dots, \boldsymbol{\theta}_{t-1}) = p(\boldsymbol{\theta}_t) = p(\boldsymbol{\theta})$ e o processo evolutivo Darwiniano $\pi(\boldsymbol{\theta}_{0:t})$ foi definido como um processo Markoviano, então $\pi(\boldsymbol{\theta}_t | \boldsymbol{\theta}_0, \dots, \boldsymbol{\theta}_{t-1}) = \pi(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1})$.

A responsabilidade da convergência de $\pi(\boldsymbol{\theta}_{0:t})$ para $p(\boldsymbol{\theta}_{0:t})$, como visto na equação anterior, reduz-se às propriedades da cadeia Markoviana $\pi(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1})$. A teoria de simulações estocásticas, garante que esta convergência só é possível caso $\pi(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1})$ seja definido como uma cadeia Markoviana *irredutível*, *aperiódica* e que satisfaça à seguinte *condição de reversibilidade* ou *condição de equilíbrio* [26]:

$$p(\boldsymbol{\theta}_t)\pi(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}) = p(\boldsymbol{\theta}_{t-1})\pi(\boldsymbol{\theta}_{t-1} | \boldsymbol{\theta}_t)\tag{10}$$

Com isto, se cumpre que:

$$\int p(\boldsymbol{\theta}_t)\pi(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1})d\boldsymbol{\theta}_t = \int p(\boldsymbol{\theta}_{t-1})\pi(\boldsymbol{\theta}_{t-1} | \boldsymbol{\theta}_t)d\boldsymbol{\theta}_t = p(\boldsymbol{\theta}_t)\tag{11}$$

mostrando que o processo Markoviano $\pi(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1})$ possui $p(\boldsymbol{\theta}_t) = p(\boldsymbol{\theta})$ como sua *distribuição estacionária*, garantindo assim a convergência do processo evolutivo Darwiniano $\pi(\boldsymbol{\theta}_{0:t})$ para o processo estocástico estacionário $p(\boldsymbol{\theta}_{0:t})$.

Na prática a definição de $\pi(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1})$ pode ser implementada como a multiplicação de duas funções [27]: Um *núcleo de transição* $g(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1})$ e uma *probabilidade de aceitação* $\alpha(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1})$, ou seja

$$\pi(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}) = g(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1})\alpha(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1})\tag{12}$$

Este *núcleo de transição* $g(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1})$, pode ser expresso através da seguinte equação matricial.

$$\begin{bmatrix} \boldsymbol{\theta}_t^{*(1)} \\ \boldsymbol{\theta}_t^{*(2)} \\ \vdots \\ \boldsymbol{\theta}_t^{*(N)} \end{bmatrix} = \begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1N} \\ q_{21} & q_{22} & \cdots & q_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ q_{N1} & q_{N2} & \cdots & q_{NN} \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta}_{t-1}^{(1)} \\ \boldsymbol{\theta}_{t-1}^{(2)} \\ \vdots \\ \boldsymbol{\theta}_{t-1}^{(N)} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_t^{(1)} \\ \boldsymbol{\varepsilon}_t^{(2)} \\ \vdots \\ \boldsymbol{\varepsilon}_t^{(N)} \end{bmatrix} \quad (13)$$

onde q_{ij} representa um escalar e $\boldsymbol{\varepsilon}_t^{(i)}$ um determinado ruído. Este núcleo não pode ser nulo, pois violaria as condições de convergência. Este núcleo pode ser a matriz identidade como expresso na seguinte equação.

$$\begin{bmatrix} \boldsymbol{\theta}_t^{*(1)} \\ \boldsymbol{\theta}_t^{*(2)} \\ \vdots \\ \boldsymbol{\theta}_t^{*(N)} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta}_{t-1}^{(1)} \\ \boldsymbol{\theta}_{t-1}^{(2)} \\ \vdots \\ \boldsymbol{\theta}_{t-1}^{(N)} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_t^{(1)} \\ \boldsymbol{\varepsilon}_t^{(2)} \\ \vdots \\ \boldsymbol{\varepsilon}_t^{(N)} \end{bmatrix} \quad (14)$$

Este caso especial é denominado de *núcleo de transição de mutação* uma vez que o cromossoma da geração atual é uma “mutação” do cromossoma da geração anterior, ou seja,

$$\boldsymbol{\theta}_t^{*(i)} = \boldsymbol{\theta}_{t-1}^{(i)} + \boldsymbol{\varepsilon}_t^{(i)} \quad (15)$$

Qualquer outro núcleo de transição, diferente do núcleo nulo e do núcleo de mutação é denominado de *núcleo de transição de cruzamento* como por exemplo,

$$\begin{bmatrix} \boldsymbol{\theta}_t^{*(i)} \\ \boldsymbol{\theta}_t^{*(j)} \end{bmatrix} = \begin{bmatrix} 1-\gamma & \gamma \\ \gamma & 1-\gamma \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta}_{t-1}^{(i)} \\ \boldsymbol{\theta}_{t-1}^{(j)} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_t^{(i)} \\ \boldsymbol{\varepsilon}_t^{(j)} \end{bmatrix} \quad (16)$$

onde $\gamma \in [0,1)$. Com relação à *probabilidade de aceitação* $\alpha(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1})$, o critério citado comumente na literatura de simulações estocásticas é o critério de Metropolis *et al.* [28] e Hastings [29] definido por:

$$\alpha(\boldsymbol{\theta}_t^* | \boldsymbol{\theta}_{t-1}) = \min\{1, \rho\} \quad (17)$$

sendo ρ a *probabilidade de sobrevivência* do cromossoma, definido por:

$$\rho = \frac{p(\boldsymbol{\theta}_t^*) \pi(\boldsymbol{\theta}_t^* | \boldsymbol{\theta}_{t-1})}{p(\boldsymbol{\theta}_t) \pi(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}^*)} \quad (18)$$

Para os cromossomas evoluídos segundo os núcleos de transição de *mutação* e de *cruzamento*, a *probabilidades de sobrevivência* ρ_m, ρ_c é dada por:

$$\rho_m = \frac{p(\boldsymbol{\theta}_t^{*(i)}) \pi(\boldsymbol{\theta}_t^{*(i)} | \boldsymbol{\theta}_t^{(i)})}{p(\boldsymbol{\theta}_t^{(i)}) \pi(\boldsymbol{\theta}_t^{(i)} | \boldsymbol{\theta}_t^{*(i)})} \quad (19)$$

$$\rho_c = \frac{p(\boldsymbol{\theta}_t^{*(i)}) p(\boldsymbol{\theta}_t^{*(j)}) \pi(\boldsymbol{\theta}_t^{*(i)}, \boldsymbol{\theta}_t^{*(j)} | \boldsymbol{\theta}_t^{(i)}, \boldsymbol{\theta}_t^{(j)})}{p(\boldsymbol{\theta}_t^{(i)}) p(\boldsymbol{\theta}_t^{(j)}) \pi(\boldsymbol{\theta}_t^{(i)}, \boldsymbol{\theta}_t^{(j)} | \boldsymbol{\theta}_t^{*(i)}, \boldsymbol{\theta}_t^{*(j)})} \quad (20)$$

À medida em que o *núcleo de transição de cruzamento* combine mais de dois cromossomas o cálculo da *probabilidade de aceitação* torna-se difícil de ser avaliada. Portanto, o critério de Metropolis e Hastings torna-se impraticável.

Para superar esta deficiência é preciso compreender que o critério de Metropolis e Hasting no fundo é um “critério de seleção” devido a que tenta eliminar os cromossomas localizados em regiões de baixa probabilidade, e simultaneamente tenta multiplicar os cromossomas localizados em regiões de alta probabilidade da distribuição $p(\theta)$. Desta forma, o critério de Metropolis e Hasting implementa na prática o princípio da “sobrevivência dos mais aptos”, só que aplicado a um número restrito de cromossomas.

Generalizando o “critério de seleção” de Metropolis e Hasting para toda a população, teremos uma versão artificial da “seleção natural” Darwiniana. Esta generalização pode ser implementada na prática através de um algoritmo computacional, como por exemplo a *seleção multinomial*, *seleção por roleta*, *seleção estocástica universal*, etc. Todos estes algoritmos são em essência esquemas de *amostragem com reposição* [20][21], onde a variável de amostragem é baseada na *aptidão normalizada* $\tilde{\omega}_i^{(t)}$ dos cromossomas.

Com base nos resultados teóricos desenvolvidos nesta seção, pode ser derivado o seguinte algoritmo computacional, denominado neste trabalho de *Filtro Evolutivo* (FE).

Procedimento Filtro Evolutivo

Início

$t \leftarrow 1$

Iniciar $\{\theta_t^{*(i)}\}_{i=1}^N$

Avaliar $\{\theta_t^{*(i)}\}_{i=1}^N$

Selecionar $\{\theta_t^{(i)}\}_{i=1}^N$ **de** $\{\theta_t^{*(i)}\}_{i=1}^N$

Enquanto (condição de parada não é satisfeita) Fazer

Início

$t \leftarrow t + 1$

Evoluir $\{\theta_t^{*(i)}\}_{i=1}^N$ **de** $\{\theta_{t-1}^{(i)}\}_{i=1}^N$

Avaliar $\{\theta_t^{*(i)}\}_{i=1}^N$

Selecionar $\{\theta_t^{(i)}\}_{i=1}^N$ **de** $\{\theta_t^{*(i)}\}_{i=1}^N$

Fim

Fim

A termo “Filtro” é devido a que o “critério de seleção” realiza uma filtragem da população, eliminando os cromossomas com aptidão baixa e multiplicando os cromossomas com aptidão alta. Os detalhes da implementação do algoritmo são explicados a seguir:

Iniciar $\{\boldsymbol{\theta}_t^{*(i)}\}_{i=1}^N$:

Para $i = 1, \dots, N$, cromossomas $\boldsymbol{\theta}_1^{*(i)} = \boldsymbol{\theta}_0^{(i)}$ sendo $\boldsymbol{\theta}_0^{(i)} \sim \pi(\boldsymbol{\theta}_0)$.

Avaliar $\{\boldsymbol{\theta}_t^{*(i)}\}_{i=1}^N$:

Para $i = 1, \dots, N$, cromossomas, avaliar a *aptidão normalizada*

$$\tilde{\omega}_t^{(i)} = \frac{p(\boldsymbol{\theta}_t^{*(i)})}{\sum_{j=1}^N p(\boldsymbol{\theta}_t^{*(j)})} \quad (21)$$

Selecionar $\{\boldsymbol{\theta}_t^{(i)}\}_{i=1}^N$ de $\{\boldsymbol{\theta}_t^{*(i)}\}_{i=1}^N$:

Amostrar com substituição N indivíduos $\{\boldsymbol{\theta}_t^{(i)}, N^{-1}\}_{i=1}^N$ a partir dos indivíduos candidatos $\{\boldsymbol{\theta}_t^{*(i)}, \tilde{\omega}_t^{(i)}\}_{i=1}^N$ de acordo com sua *aptidão normalizada*.

Evoluir $\{\boldsymbol{\theta}_t^{*(i)}\}_{i=1}^N$ de $\{\boldsymbol{\theta}_{t-1}^{(i)}\}_{i=1}^N$:

Para $i = 1, \dots, N/2$, $j = N - i + 1$, $u \sim \mathbf{U}(0,1)$

$$(\boldsymbol{\theta}_t^{(i)}, \boldsymbol{\theta}_t^{(j)}) = \begin{cases} \sim \pi(\cdot | \boldsymbol{\theta}_{t-1}^{(i)}, \boldsymbol{\theta}_{t-1}^{(j)}) & \text{se } u \leq \rho_c \\ (\boldsymbol{\theta}_{t-1}^{(i)}, \boldsymbol{\theta}_{t-1}^{(j)}) & \text{caso contrário} \end{cases} \quad (22)$$

Para $i = 1, \dots, N$, $u \sim \mathbf{U}(0,1)$

$$\boldsymbol{\theta}_t^{*(i)} = \begin{cases} \sim \pi(\cdot | \boldsymbol{\theta}_t^{(i)}) & \text{se } u \leq \rho_m \\ \boldsymbol{\theta}_t^{(i)} & \text{caso contrário} \end{cases} \quad (23)$$

Fixar $\boldsymbol{\theta}_{0:t}^{*(i)} = \{\boldsymbol{\theta}_{0:t-1}^{(i)}, \boldsymbol{\theta}_t^{*(i)}\}$.

Note que na equação (21) não aparece $\tilde{\omega}_{t-1}^{(i)}$ uma vez que os indivíduos a serem evoluídos $\{\boldsymbol{\theta}_{0:t-1}^{(i)}; i = 1, \dots, N\}$ possuem pesos uniformes após o estágio *Selecionar*, no tempo $t - 1$. A cada geração t do algoritmo sempre teremos um cromossoma mais apto, conhecido como cromossoma elitista $\boldsymbol{\theta}_t^{\text{elitista}}$. Então

aplicando a lei dos grandes números $N \rightarrow \infty$ e $t \rightarrow \infty$, é esperado que este cromossoma elitista seja exatamente a moda global θ^{\max} de $p(\theta)$ e, como $\theta^{\max} \equiv \theta^{\min}$, então este mesmo ponto é o mínimo valor global da função $E(\theta)$.

Logo se conclui que:

$$\theta^{\text{elitista}} = \theta^{\max} = \theta^{\min} \quad (24)$$

Provando desta maneira a convergência do processo evolutivo para $\theta^{\max} \equiv \theta^{\min}$ e simultaneamente para a distribuição $p(\theta)$.

Para o exemplo de otimização descrito na Seção 2.3 as figuras (6) e (7) mostram a evolução dos cromossomas ao longo do tempo, tanto para a distribuição marginal de $p(x_t)$ quando para a distribuição marginal $p(y_t)$. Note que à medida que transcorre o tempo, os cromossomas localizam-se nas regiões de maior probabilidade da distribuição e conseqüentemente da sua moda global.

Já na figura (8) mostra-se a evolução do cromossoma elitista, o cromossoma de maior aptidão ao longo das gerações, nele pode-se notar sua convergência à moda global da distribuição $p(\theta_t) = p(x_t, y_t)$. Na seção seguinte será mostrada a equivalência entre o princípio Darwiniano e o teorema de Bayes.

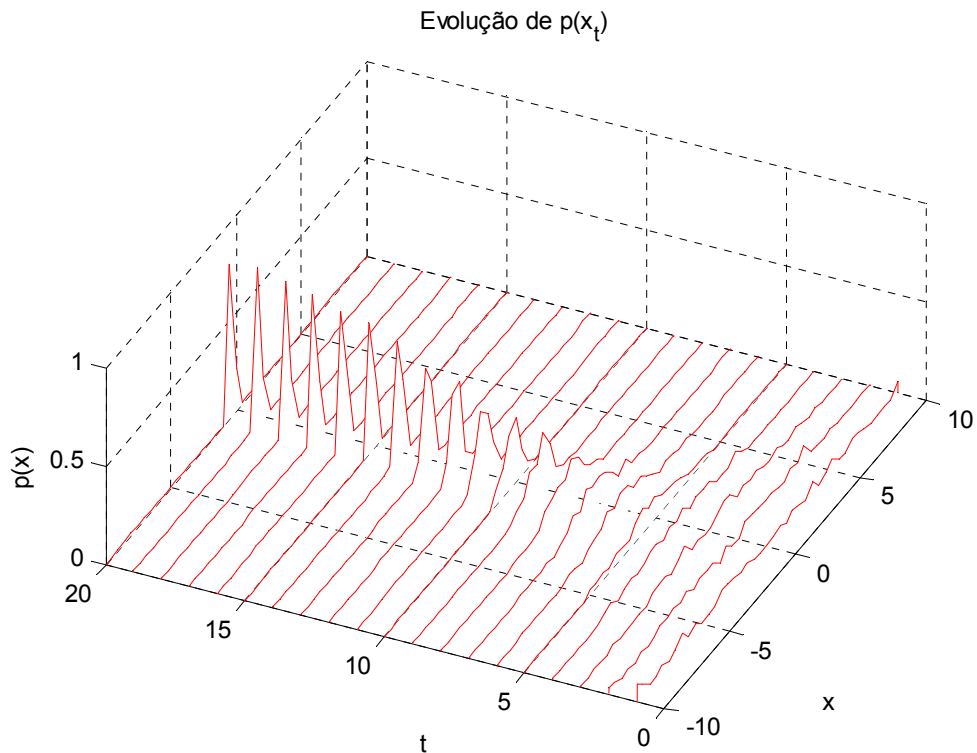


Figura 6. Evolução da distribuição marginal $p(x_t)$.

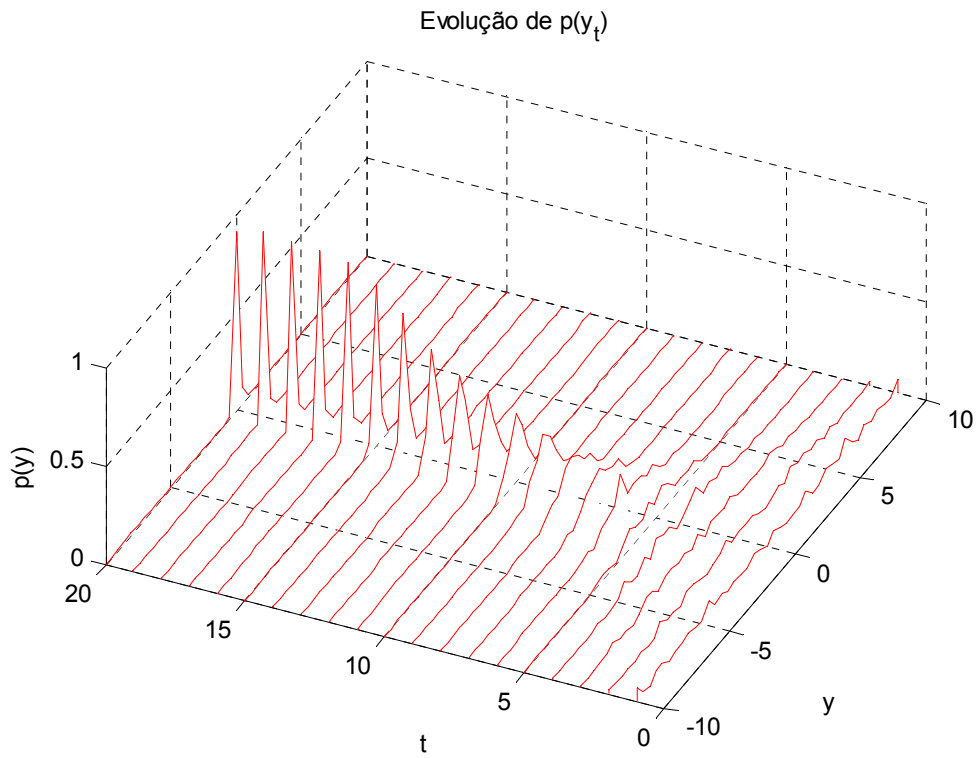


Figura 7. Evolução da distribuição marginal $p(y_t)$.

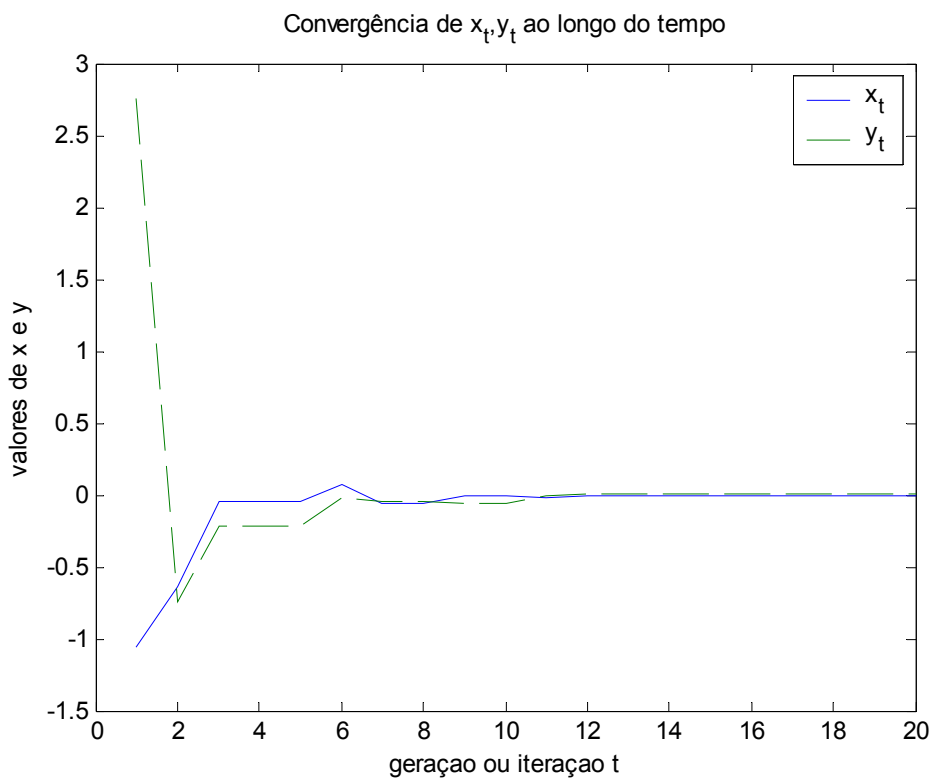


Figura 8. Convergência da moda da distribuição $p(\theta_t) = p(x_t, y_t)$.

2.3.2.

Equivalência Entre o Princípio Darwiniano e o Teorema de Bayes

Em muitas aplicações reais de processamento de sinais, como é o caso do sistema de visão artificial, as imagens digitais chegam seqüencialmente no tempo. Esta seqüência de imagens digitais podem ser modeladas por uma seqüência de variáveis aleatórias $\mathbf{y}_{1:t} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_t\}$ definidas num espaço m-dimensional $\mathbf{y}_i \in \mathfrak{R}^m$ (detalhes desta notação vetorial encontra-se na próximo capítulo).

Por outro lado, qualquer modelo de processamento de sinais pode ser expresso de forma genérica através de um Modelo de Espaço de Estados (MEE). Este modelo abstrai o modelo original através de duas novas equações: a *equação de estado* e a *equação de observação*. Estas equações podem ser descritas probabilisticamente através das seguintes expressões matemáticas:

$$p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}) \quad (25)$$

$$p(\mathbf{y}_t | \boldsymbol{\theta}_t) \quad (26)$$

sendo $\boldsymbol{\theta} \in \mathfrak{R}^n$ os “parâmetros adaptativos”, ou “estados” do MEE. A *equação de estado* $p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1})$ é tipicamente modelada como um processo estocástico Markoviano, ou seja, $p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_{t-2}, \dots, \boldsymbol{\theta}_0) = p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1})$, e a *equação de observação* $p(\mathbf{y}_t | \boldsymbol{\theta}_t)$ é modelada de forma que as observações sejam condicionalmente independentes aos estados $\boldsymbol{\theta}_t$.

Neste contexto, a tarefa de análise dos dados destas imagens digitais envolve estimar os estados $\boldsymbol{\theta}_{0:t} = \{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_t\}$ a cada imagem \mathbf{y}_i observada do sinal de vídeo $\mathbf{y}_{1:t} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_t\}$. Nestas aplicações de processamento de sinais, sempre é possível ter algum conhecimento prévio sobre os sinais que estão sendo modelados. Este conhecimento prévio nos permite formular modelos Bayesianos [30][31], isto é, uma distribuição *a priori* sobre os estados $p(\boldsymbol{\theta}_{0:t})$ e uma função de verossimilhança $p(\mathbf{y}_{1:t} | \boldsymbol{\theta}_{0:t})$ que relaciona estes estados às imagens observadas no sinal de vídeo. Dentro desta configuração, a inferência Bayesiana dos estados é baseada na distribuição *a posteriori* $p(\boldsymbol{\theta}_{0:t} | \mathbf{y}_{1:t})$, que é obtida através do teorema de Bayes, expresso pela seguinte equação:

$$p(\boldsymbol{\theta}_{0:t} | \mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_{1:t} | \boldsymbol{\theta}_{0:t})}{p(\mathbf{y}_{1:t})} p(\boldsymbol{\theta}_{0:t}) \quad (27)$$

As distribuições envolvidas na equação anterior são conhecidas como:

$$Posteriori = \frac{Verossimilhança}{Preditiva} Priori$$

A distribuição *preditiva* é um fator de normalização que garante que a integral $\int p(\boldsymbol{\theta}_{0:t} | \mathbf{y}_{1:t}) d\boldsymbol{\theta}_{0:t}$ seja igual à unidade.

Embora $p(\boldsymbol{\theta}_{0:t} | \mathbf{y}_{1:t})$ constitua a inferência completa dos estados do MEE, tipicamente, é de interesse calcular a cada instante de tempo t sua correspondente distribuição marginal $p(\boldsymbol{\theta}_t | \mathbf{y}_{1:t})$ denominada também como *distribuição de filtragem*. Contudo, a forma mais eficiente de calcular esta *distribuição de filtragem* é através de um cálculo recursivo. Com estes propósitos a equação (27) pode ser expressa equivalentemente através das equações de “atualização” e de “previsão” mostradas a seguir [32]-[40].

$$p(\boldsymbol{\theta}_t | \mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_t | \boldsymbol{\theta}_t)}{p(\mathbf{y}_{1:t})} p(\boldsymbol{\theta}_t | \mathbf{y}_{1:t-1}) \quad (28)$$

$$p(\boldsymbol{\theta}_t | \mathbf{y}_{1:t-1}) = \int p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}) p(\boldsymbol{\theta}_{t-1} | \mathbf{y}_{1:t-1}) d\boldsymbol{\theta}_{t-1} \quad (29)$$

As equações de atualização e de previsão, envolvem integrais multidimensionais difíceis de serem tratadas analiticamente. Uma forma de contornar este problema é calculando esta distribuição *a posteriori* numericamente, através de um processo de simulações estocásticas.

Neste preciso ponto, estamos diante de um problema de “amostragem” de uma distribuição de probabilidade, a qual foi tratada amplamente na seção anterior. Só que desta vez a distribuição objetivo $p(\boldsymbol{\theta}_t | \mathbf{y}_{1:t})$ é um processo estocástico dinâmico, uma vez que os estados mudam em função dos dados observados. Porém o processo evolutivo Darwiniano, exposto na seção anterior, precisa ser adaptado a para estas novas condições. Para começar, o *peso de importância* equivalente à equação (4) estaria formulado como [32]-[40]:

$$\begin{aligned} \omega(\boldsymbol{\theta}_{0:t}) &= \frac{p(\boldsymbol{\theta}_{0:t} | \mathbf{y}_{1:t})}{\pi(\boldsymbol{\theta}_{0:t} | \mathbf{y}_{1:t})} \\ &= \omega(\boldsymbol{\theta}_{0:t-1}) \frac{p(\boldsymbol{\theta}_{0:t} | \mathbf{y}_{1:t})}{\pi(\boldsymbol{\theta}_{0:t} | \mathbf{y}_{1:t})} \frac{\pi(\boldsymbol{\theta}_{0:t-1} | \mathbf{y}_{1:t-1})}{p(\boldsymbol{\theta}_{0:t-1} | \mathbf{y}_{1:t-1})} \end{aligned} \quad (30)$$

Aplicando o teorema de Bayes em $p(\boldsymbol{\theta}_{0:t} | \mathbf{y}_{1:t})$ e $p(\boldsymbol{\theta}_{0:t-1} | \mathbf{y}_{1:t-1})$ obtém-se:

$$\omega(\boldsymbol{\theta}_{0:t}) = \omega(\boldsymbol{\theta}_{0:t-1}) \frac{p(\mathbf{y}_{1:t} | \boldsymbol{\theta}_{0:t})}{p(\mathbf{y}_{1:t-1} | \boldsymbol{\theta}_{0:t-1})} \frac{p(\boldsymbol{\theta}_{0:t})}{p(\boldsymbol{\theta}_{0:t-1})} \frac{p(\mathbf{y}_{1:t-1})}{p(\mathbf{y}_{1:t})} \frac{\pi(\boldsymbol{\theta}_{0:t-1} | \mathbf{y}_{1:t-1})}{\pi(\boldsymbol{\theta}_{0:t} | \mathbf{y}_{1:t})} \quad (31)$$

Uma vez que os estados seguem um processo Markoviano, então:

$$p(\boldsymbol{\Theta}_{0:t}) = p(\boldsymbol{\theta}_0) \prod_{i=1}^t p(\boldsymbol{\theta}_i | \boldsymbol{\theta}_{i-1}) \quad (32)$$

Como as observações são condicionalmente independentes dos estados, temos:

$$p(\mathbf{Y}_{1:t} | \boldsymbol{\Theta}_{0:t}) = \prod_{i=1}^t p(\mathbf{y}_i | \boldsymbol{\theta}_i) \quad (33)$$

Todavia, a fim de aproveitar os cromossomas da população da geração anterior $\{\boldsymbol{\theta}_{0:t-1}^{(i)}; i=1, \dots, N\}$ de $p(\boldsymbol{\theta}_{0:t-1} | \mathbf{y}_{1:t-1})$, no tempo $t-1$, para o cálculo de $p(\boldsymbol{\theta}_{0:t} | \mathbf{y}_{1:t})$, no tempo t , é necessário que o processo evolutivo Darwiniano $\pi(\boldsymbol{\theta}_{0:t} | \mathbf{y}_{1:t})$, no tempo t , admita como densidade marginal no tempo $t-1$ a expressão $\pi(\boldsymbol{\theta}_{0:t-1} | \mathbf{y}_{1:t-1})$, ou seja:

$$\pi(\boldsymbol{\theta}_{0:t} | \mathbf{y}_{1:t}) = \pi(\boldsymbol{\theta}_{0:t-1} | \mathbf{y}_{1:t-1}) \pi(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{0:t-1}, \mathbf{y}_{1:t}) \quad (34)$$

Substituindo adequadamente as equações (32), (33) e (34) em (31) obtém-se:

$$\omega(\boldsymbol{\theta}_{0:t}) = \omega(\boldsymbol{\theta}_{0:t-1}) \frac{p(\mathbf{y}_t | \boldsymbol{\theta}_t) p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1})}{\pi(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{0:t-1}, \mathbf{y}_{1:t})} \frac{p(\mathbf{y}_{1:t-1})}{p(\mathbf{y}_{1:t})} \quad (35)$$

Finalmente, o *peso de importância normalizado* $\tilde{\omega}_t^{(i)}$, conhecido também como *aptidão normalizada*, para este problema seria expresso por:

$$\begin{aligned} \tilde{\omega}_t^{(i)} &= \frac{\omega(\boldsymbol{\theta}_{0:t}^{(i)})}{\sum_{j=1}^N \omega(\boldsymbol{\theta}_{0:t}^{(j)})} \\ &\propto \tilde{\omega}_{t-1}^{(i)} \frac{p(\mathbf{y}_t | \boldsymbol{\theta}_t^{(i)}) p(\boldsymbol{\theta}_t^{(i)} | \boldsymbol{\theta}_{t-1}^{(i)})}{\pi(\boldsymbol{\theta}_t^{(i)} | \boldsymbol{\theta}_{0:t-1}^{(i)}, \mathbf{y}_{1:t})} \end{aligned} \quad (36)$$

Na equação anterior $p(\mathbf{y}_t | \boldsymbol{\theta}_t^{(i)})$ é fácil de ser avaliada pois depende exclusivamente da equação de observação, a qual é uma equação conhecida do MEE. Analogamente $p(\boldsymbol{\theta}_t^{(i)} | \boldsymbol{\theta}_{t-1}^{(i)})$ é fácil de ser avaliada, pois depende exclusivamente da equação de estado, que também é uma equação conhecida do MEE. Entretanto $\pi(\boldsymbol{\theta}_t^{(i)} | \boldsymbol{\theta}_{0:t-1}^{(i)}, \mathbf{y}_{1:t})$ é o processo evolutivo Darwiniano que precisa ser definido cuidadosamente para garantir a convergência de $\pi(\boldsymbol{\theta}_{0:t} | \mathbf{y}_{1:t})$ para $p(\boldsymbol{\theta}_{0:t} | \mathbf{y}_{1:t})$.

Da Seção 2.3.1 sabemos que a cadeia Markoviana $\pi(\boldsymbol{\theta}_t^{(i)} | \boldsymbol{\theta}_{0:t-1}^{(i)}, \mathbf{y}_{1:t})$ tem que ser *irredutível*, *aperiódica* e satisfazer *condição de reversibilidade* ou *condição de equilíbrio*. Além destas condições, a cadeia Markoviana tem que acompanhar a dinâmica dos dados de observação. Para que isto seja possível, é necessário que o núcleo de transição de *mutação* e/ou de *cruzamento* possuam uma dinâmica “equivalente” à dinâmica dos estados (definido pela equação de estado) ou seja $\pi(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{0:t-1}, \mathbf{y}_{1:t}) \equiv p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1})$. Portanto:

$$\tilde{\omega}_t^{(i)} \propto \tilde{\omega}_{t-1}^{(i)} p(\mathbf{y}_t | \boldsymbol{\theta}_t^{(i)}) \quad (37)$$

Com base no pseudocódigo do Filtro Evolutivo descrito anteriormente, a equação (37) estaria localizada no estágio “Avaliar”. Simultaneamente, esta equação possui uma relação diretamente com a distribuição de verossimilhança $p(\mathbf{y}_t | \boldsymbol{\theta}_t)$ da equação (28) na qual a distribuição *a priori* $p(\boldsymbol{\theta}_t | \mathbf{y}_{1:t-1})$ está relacionada com os estágios “Iniciar” e “Evoluir”, já que para $t = 0$ o processo evolutivo Darwiniano inicia-se com a mesma distribuição *a priori* dos estados $\pi(\boldsymbol{\theta}_0) = p(\boldsymbol{\theta}_0)$, e para $t > 0$ o processo evolutivo Darwiniano $\pi(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{0:t-1}, \mathbf{y}_{1:t})$ evolui com a mesma dinâmica da equação de estados $p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1})$ a qual está imersa indiretamente em $p(\boldsymbol{\theta}_t | \mathbf{y}_{1:t-1})$ através da equação (29). Finalmente a distribuição *a posteriori* $p(\boldsymbol{\theta}_t | \mathbf{y}_{1:t})$ da equação (28) está relacionada com o estágio “Selecionar”, uma vez que este estágio, derivado do critério de Metropolis e Hasting, tem a missão de selar a *condição de reversibilidade* ou *condição de equilíbrio* da cadeia Markoviana $\pi(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{0:t-1}, \mathbf{y}_{1:t})$ de forma que $p(\boldsymbol{\theta}_t | \mathbf{y}_{1:t})$ seja sua *distribuição estacionária*.

Com base nestes argumentos pode-se concluir a existência de uma equivalência entre o princípio Darwiniano da “Sobrevivência dos mais aptos” e o teorema de Bayes, como mostrado no seguinte gráfico.

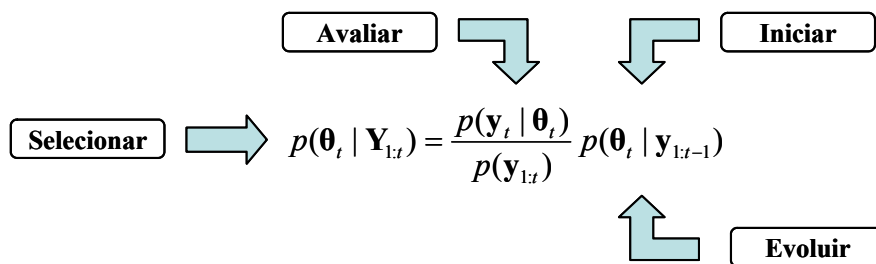


Figura 9. Equivalência entre o princípio Darwiniano e o teorema de Bayes.

Com base neste resultado, o problema de visão artificial pode ser analisado do ponto de vista Bayesiano, pois de forma indireta estaríamos tratando o problema numa visão Darwiniana. Além disso, tem-se a vantagem que a modelagem Bayesiana em certos casos pode fornecer resultados analíticos fechados que podem ser aproveitados para minimizar o esforço computacional. Para os casos em que a solução não admita um tratamento analítico fechado, a solução será através de algoritmos numéricos abordados do ponto de vista da “sobrevivência dos mais aptos”.

2.4. Resumo

Neste capítulo foi apresentada à natureza da visão humana. Basicamente, foram utilizadas a teoria da seleção natural dos replicadores e a teoria computacional da mente para entender o processo evolutivo que criou nosso corpo e nossa mente.

Com base na generalidade do princípio Darwiniano, foi proposta uma extensão aplicada a problemas de amostragem e/ou otimização de uma distribuição de probabilidades “incondicional” $p(\theta_{0:t})$, caracterizada por ser estacionária e invariante no tempo.

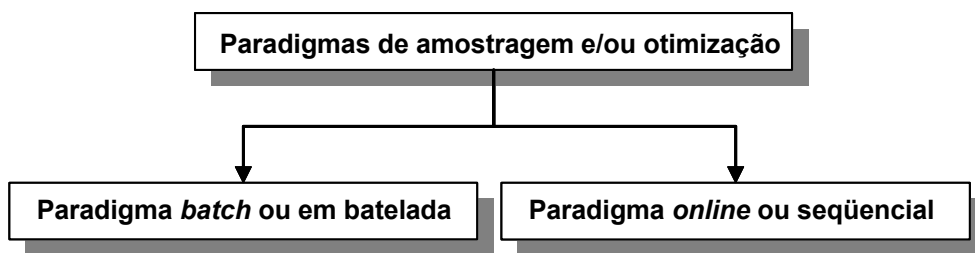


Figura 10. Paradigmas de amostragem e/ou otimização.

Entretanto, freqüentemente precisa-se amostrar e/ou otimizar uma distribuição de probabilidades $p(\theta_{0:t} | \mathbf{y}_{1:t})$ “condicional” a um conjunto de observações $\mathbf{y}_{1:t} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_t\}$; $\mathbf{y}_i \in \mathfrak{R}^m$. Neste ponto surgem dois paradigmas de amostragem e/ou otimização (figura 10): o paradigma *batch* ou em batelada, onde a amostragem e/ou otimização é realizada sobre todo o conjunto de

observações simultaneamente; e o paradigma *online* ou seqüencial, onde a amostragem e/ou otimização é realizada seqüencialmente no tempo por cada observação concebida do processo.

A amostragem e/ou otimização em batelada de $p(\boldsymbol{\theta}_{0:t} | \mathbf{y}_{1:t})$ tem o mesmo tratamento explicado na Seção 2.3.1. Para o caso da amostragem e/ou otimização seqüencial de $p(\boldsymbol{\theta}_{0:t} | \mathbf{y}_{1:t})$, a Seção 2.3.2 mostra os detalhes deste processamento.

A escolha do paradigma de amostragem e/ou otimização depende da aplicação desenvolvida. Por exemplo, para o caso de visão computacional, poderia ser utilizado um amostrador/otimizador em batelada para “treinar” de forma “*off-line*” o sistema de visão artificial, e em seguida poderia ser utilizado um amostrador/otimizador seqüencial para continuar ajustando de forma “*on-line*” o sistema de visão artificial. Além disso, poderia ser utilizado um outro amostrador/otimizador seqüencial encarregado de ajustar o modelo de seguimento ou “*tracking*” do objeto.

Note que no próximo capítulo a solução do modelo Bayesiano de visão artificial foi calculado utilizando somente amostradores/otimizadores em batelada.

3 Visão Artificial por Métodos Computacionais

A visão humana começa quando um fóton (unidade de energia luminosa) reflete-se de uma superfície e atravessa rapidamente a pupila por uma linha, para estimular um dos fotorreceptores (bastonetes e cones) que revestem a superfície interna curva do globo ocular (retina). O receptor transmite um sinal neural ao cérebro, e a primeira tarefa do cérebro é descobrir de que parte do mundo veio esse fóton. Infelizmente, o raio que define a trajetória do fóton estende-se ao infinito, e tudo o que o cérebro sabe é que o retalho que o originou encontra-se em algum lugar ao longo do raio. A informação sobre a distância da profundidade do retalho ao olho, perdeu-se no processo de projeção. A ambigüidade é multiplicada de modo combinatório pelos milhões de outros receptores da retina. Qualquer imagem retiniana, portanto, poderia ter sido produzida por um número infinito de arranjos de superfícies tridimensionais no mundo (Fig. 11).

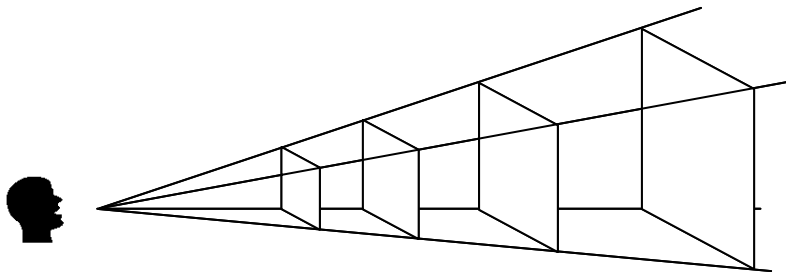


Figura 11. Problema da profundidade.

Obviamente, não percebemos infinitas possibilidades; olhamos uma, geralmente próxima à correta. A evolução solucionou este problema através da visão estereoscópica na qual são capturadas duas projeções ligeiramente diferentes, uma em cada olho, e a partir delas o cérebro executa cálculos trigonométrico para reconstruir a terceira dimensão.

Entretanto, a dedução da forma de um objeto – reconhecimento de objetos – a partir da sua projeção, é um “problema mal proposto” pois é um problema que não tem uma solução única. Uma forma elíptica na retina poderia ter provindo de uma oval vista de frente ou de um círculo visto obliquamente. A visão evoluiu de

modo a converter esses problemas mal propostos em problemas solúveis adicionando “premissas” [15]: suposições sobre como, em média, o mundo em que evoluímos é montado. Por exemplo o sistema visual humano “supõe” que a matéria é coesa, as superfícies são uniformemente coloridas entre outras. Quando estas suposições são violadas, somos presas de ilusões. O exemplo é o triângulo de Kanisza (figura 12), um trecho de nada que delinea uma forma tão real como se ela estivesse desenhada a tinta.

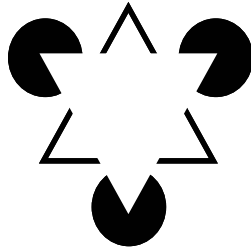


Figura 12. Ilusão óptica: Triângulo de Kanisza.

Seguindo com este raciocínio de adição de “premissas”, além das idéias fornecidas pela teoria de seleção natural de replicadores, a teoria computacional da mente, na próxima seção será desenvolvido o modelo de visão computacional proposto neste trabalho.

3.1. Modelo Probabilístico de Visão Artificial

Imitar artificialmente todas as habilidades da visão humana é uma tarefa extremamente complexa. Porém há necessidade de concentrar os esforços na replicação de algumas habilidades. Este trabalho, em particular, concentrou seus esforços no estudo do “reconhecimento de objetos tridimensionais em ambientes dinâmicos”, como o exemplo da figura (13).

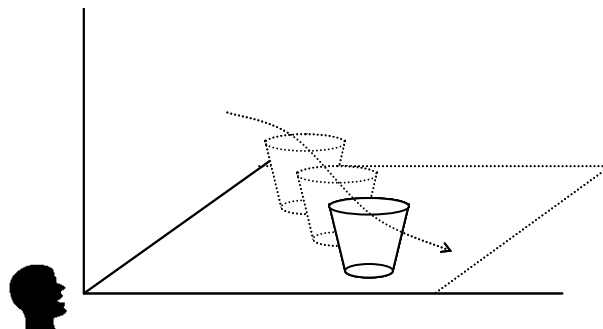


Figura 13. Reconhecimento de objetos tridimensionais em ambientes dinâmicos.

imagens. Por esta razão, a comunidade de visão artificial, tipicamente, usa uma notação vetorial. Assim qualquer imagen digital matricial de $(q \times r)$ pixels, pode ser expreso por um vetor m -dimensional, como ilustrado na seguinte figura.

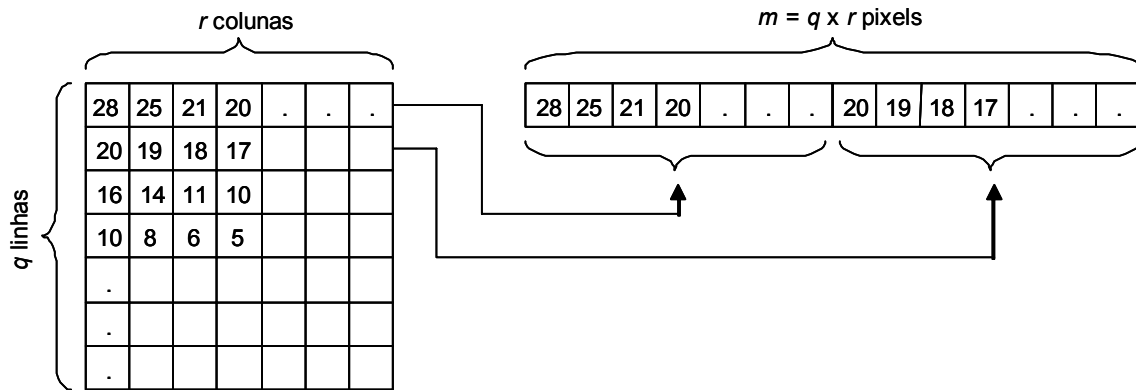


Figura 15. Representação vetorial de uma imagem.

Sob esta nova perspectiva, qualquer imagem digital, pode ser vista como um “ponto” imerso num espaço m -dimensional. Portanto, é razoável supor como “premissa” que as infinitas vistas de um objeto tridimensional formam uma “curva contínua” nesse espaço dimensional, como ilustrado na figura (16). Evidentemente, por questões praticas, podemos selecionar um número finito de vistas representativas do objeto as quais podem ser consideradas os “centróides” de um agrupamento de pontos, como ilustrado nessa mesma figura.

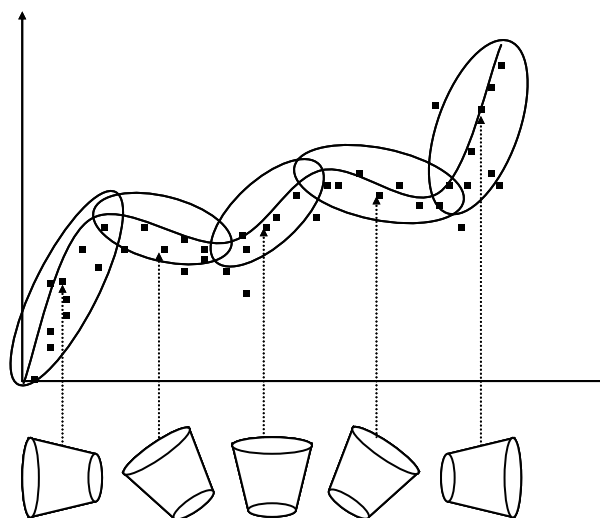


Figura 16. Curva formada pelas infinitas vistas do objeto.

Portanto, de maneira genérica, o sistema de visão artificial tem como objetivo, primeiramente, “aprender a função contínua” com base nos agrupamentos de vistas selecionadas; e em segundo lugar, “aprender a dinâmica do objeto” gerada pelo deslocamento e rotação do objeto que mudam a vista do objeto.

Estas duas tarefas aparentemente simples, tornam-se complicadas devido à “explosão de dimensionalidade” dos dados processados. Por exemplo, para um vídeo formado por imagens de (64x64) pixels, o espaço dimensional dos dados é de 4096 componentes. Conseqüentemente, não é possível processar os sinais de vídeo diretamente no espaço dimensional real, como estudado na Seção 2.3.2. Porém, surge naturalmente a necessidade de modularizar o sistema de visão artificial. Como visto no capítulo inicial, a comunidade de Inteligência Artificial, tipicamente, modulariza estes sistemas em três módulos [41]: (a) aquisição de dados, (b) pré-processamento de dados e (c) classificação de padrões. Para o sistema de visão artificial, proposto neste trabalho (figura 17), o módulo de aquisição de dados estaria composto por um Sistema Mono Ocular (SMO) com filtros digitais destinados a melhorar a qualidade da imagem; o módulo de pré-processamento de dados estaria composto por um método de redução de dimensionalidade, como por exemplo a Análise de Componentes Principais (ACP); e o módulo de classificação de padrões estaria composto por Redes Neurais Artificiais (RNA).

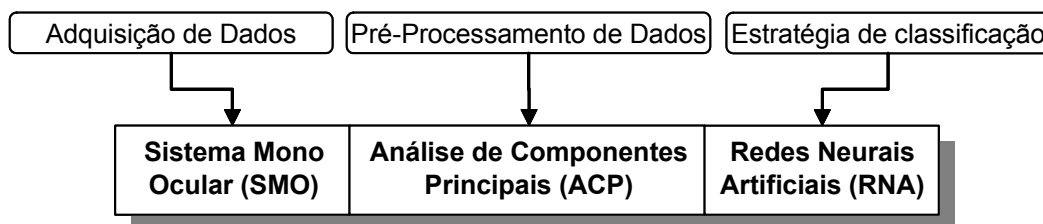


Figura 17. Modelo de visão computacional proposto neste trabalho.

O fluxo dos sinais através dos módulos do sistema artificial (figura 18) começaria com a digitalização do sinal luminoso real $s_t \in \mathfrak{R}^\infty$ através do SMO e como saída deste módulo teríamos o sinal $y_t \in \mathfrak{R}^m$ o qual é um sinal equalizado e filtrado de ruído. Em seguida o sinal y_t que apresenta um número elevado de componentes dimensionais é transformado pelo módulo de ACP para um sinal

equivalente $\mathbf{x}_t \in \mathfrak{R}^d$ mas que apresenta um número muito menor de componentes dimensionais. Finalmente, o sinal \mathbf{x}_t é discriminado pela módulo de RNA gerando como resultado o sinal $\mathbf{z}_t \in \mathfrak{R}^c$ sendo c a número de vistas ou agrupamentos interesse.

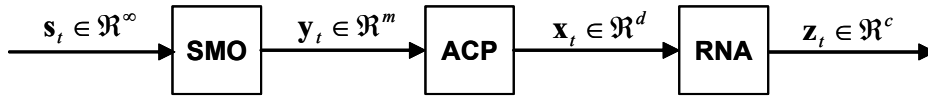


Figura 18. Fluxo dos sinais através dos módulos do sistema.

De forma genérica, cada módulo do sistema de visão artificial, pode ser visto como uma transformação matemática (equações paramétricas) que mapeia ao longo do tempo pontos entre espaços dimensionais diferentes. Assim, todo o Sistema de Visão Artificial (SVA) pode ser visto como um modelo paramétrico (figura 19). Sob esta ótica, o problema de visão artificial torna-se num problema de estimação seqüencial de parâmetros por cada imagem processada no sistema.

Entretanto, devido ao fato de que o objeto tridimensional é um objeto com estrutura física fixa, os parâmetros do modelo tendem a ter um valor estacionário ao longo do tempo. Isto permite evitar o cálculo seqüencial de parâmetros para um único cálculo de parâmetros realizado de forma simultânea sobre todas as imagens do objeto $\mathbf{y}_{1:t} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_t\}$.

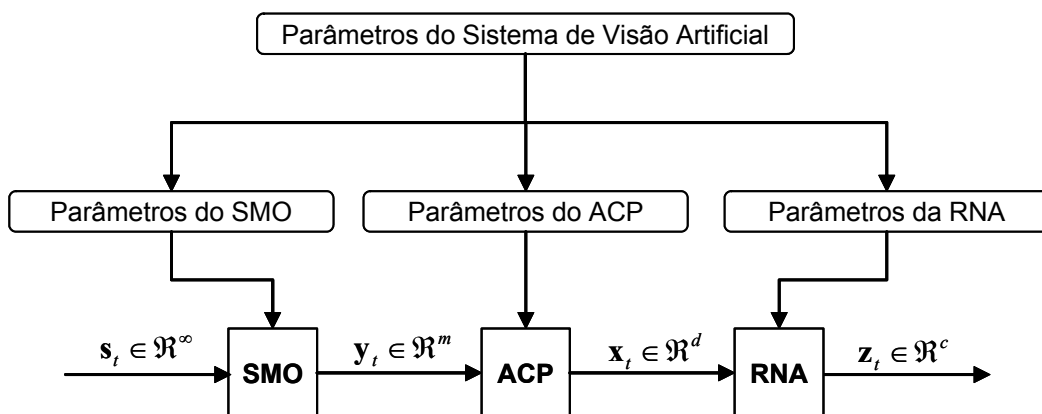


Figura 19. Modelo paramétrico de sistema de visão artificial.

A seguir será discutida a estimação de parâmetros módulo a módulo e posteriormente será apresentado uma estimação global para todo o sistema.

3.1.1. Módulo do Sistema Mono-Ocular

O módulo do SMO basicamente tem por objetivo capturar os sinais luminosos do mundo exterior $s_t \in \mathfrak{R}^\infty$ numa imagem digital $y_t \in \mathfrak{R}^m$. Como parte deste processo de captura e digitalização, podem ser incorporados internamente filtros de remoção de ruído, equalização de histograma, entre outros. A escolha dos filtros esta intimamente ligada às características específicas da aplicação, por esta razão não entraremos nestes detalhes.

O relevante neste módulo, para os propósitos deste trabalho, é que o sensor ótico assim como os filtros possuem parâmetros (figura 20) que precisam ser calibrados para alcançar um ótimo funcionamento do módulo.

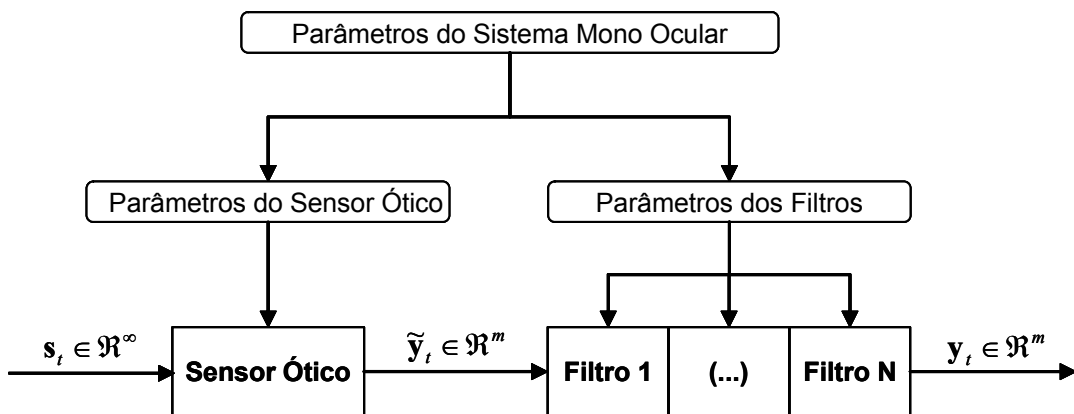


Figura 20. Parâmetros do Sistema Mono Ocular.

Entretanto, esta calibração, tipicamente, é realizado manualmente de acordo com certos critérios do projetista do sistema. Evidentemente, este processo manual representa um ponto crítico do sistema de visão computacional. Para este problema em particular, este trabalho propõe uma solução a qual será apresentada na parte final deste capítulo.

A próxima seção trata do módulo de APC encarregado basicamente de diminuir a dimensão das imagens do sistema de visão artificial.

3.1.2. Módulo da Análise de Componentes Principais

Considere uma seqüência de imagens $\mathbf{y}_{1:t} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_k, \dots, \mathbf{y}_t\}$; $\mathbf{y}_k \in \mathfrak{R}^m$; obtidas após a fase de aquisição de dados. Estas imagens, como vistos anteriormente, formam uma nuvem de pontos no espaço m-dimensional.

Uma forma de analisar a concentração e dispersão destes pontos, nesse espaço dimensional, é através das estatísticas da média e da matriz de covariância, estimadas respectivamente pelas seguintes equações:

$$\bar{\mathbf{y}} = \frac{1}{t} \sum_{k=1}^t \mathbf{y}_k \quad (38)$$

$$\mathbf{S} = \frac{1}{t} \sum_{k=1}^t (\mathbf{y}_k - \bar{\mathbf{y}})(\mathbf{y}_k - \bar{\mathbf{y}})^T \quad (39)$$

Com base nestas métricas de concentração e dispersão é possível definir uma elipse tal que englobe todas a nuvem de pontos. A equação desta elipse é dada por:

$$(\mathbf{y} - \bar{\mathbf{y}})^T \mathbf{S}^{-1} (\mathbf{y} - \bar{\mathbf{y}}) = \mathbf{r}^2 \quad (40)$$

A figura (20) mostra um exemplo ilustrativo da concentração e dispersão das imagens para ($m = 2$). Além disso, é mostrada uma elipse que engloba toda a nuvem de pontos. Neste mesmo exemplo, pode-se observar também que a elipse genericamente não é canônica, ou seja a elipse possui uma inclinação e seu centro não coincide com a origem de coordenadas.

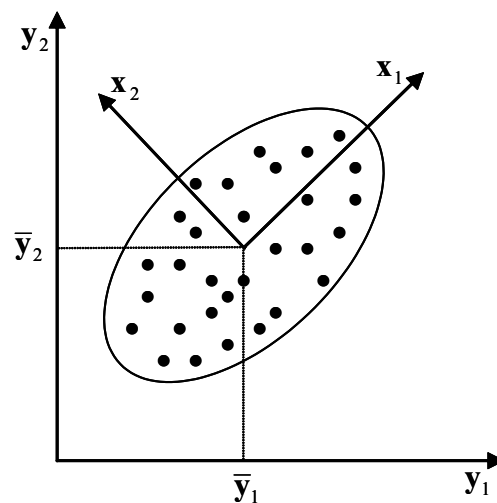


Figura 21. Concentração e dispersão das imagens.

Entretanto, sempre é possível transformar uma elipse não-canônica em canônica, realizando uma translação do eixo de coordenadas para o centro da elipse e simultaneamente uma rotação dos eixos, como mostrado na figura (21).

A idéia básica de redução de dimensionalidade está no fato de que determinados raios da elipse, no novo eixo de coordenadas, podem apresentar valores numéricos muito menores com relação a os outros raios, por exemplo, na figura (22) o raio no eixo x_2 é muito menor que o raio do eixo x_1 .

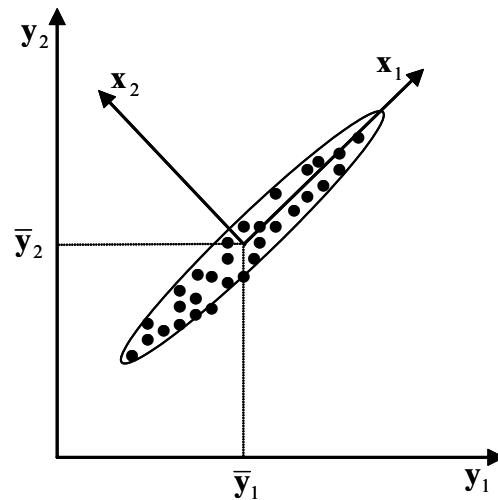


Figura 22. Redução de dimensionalidade (eliminação do eixo x_2).

Isto significa, de maneira genérica, que há pouca dispersão da nuvem de pontos nestes eixos da elipse. Portanto, estes eixos podem ser suprimidos sem comprometer a informação global da concentração e dispersão da nuvem de pontos. Como resultado desta eliminação de eixos, temos uma redução da dimensionalidade dos dados.

Para executar esta redução de dimensionalidade é necessário transformar a elipse não-canônica em canônica realizando, inicialmente, uma translação da origem do eixo de coordenadas para o centro da elipse (equação 38); seguir-se-á de uma rotação do eixo de coordenadas. Sabe-se que os raios da elipse assim como as direções do novo eixo de coordenadas estão relacionados, com os autovalores λ_i e vetores próprios \mathbf{u}_i da matriz de covariância \mathbf{S} (equação 39) os quais são calculados através da seguinte equação [42][43]:

$$\mathbf{S}\mathbf{u}_i = \lambda_i\mathbf{u}_i \tag{41}$$

Sendo $i = 1, 2, \dots, m$. Desta forma, os vetores próprios correspondentes aos d maiores autovalores (sendo $d < m$) são retidos, e a nuvem de pontos neste novo espaço dimensional reduzido é definida por:

$$\mathbf{x}_k = \mathbf{U}^T (\mathbf{y}_k - \bar{\mathbf{y}}) \quad (42)$$

Sendo $k = 1, 2, \dots, t$ e $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_d)$. Finalmente, todo o processo de redução de dimensionalidade requer apenas um parâmetro que é o número de componentes principais d , como mostrado na seguinte figura.

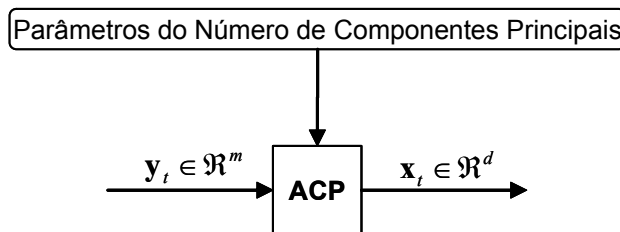


Figura 23. Parâmetros do módulo de redução de dimensionalidade.

3.1.3. Módulo de Redes Neurais Artificiais

As Redes Neurais Artificiais (RNA) [44]-[47] são modelos paramétricos que possuem a capacidade de aproximar qualquer função matemática. A arquitetura mais conhecida é a arquitetura Perceptron com Múltiplas Camadas (PMC). A mais utilizada é a arquitetura com duas camadas (figura 24).

Basicamente, a rede neural mapeia pontos entre um espaço d -dimensional e um c -dimensional. Para isto, é necessário ter um conjunto de “exemplos deste mapeamento”, os quais são utilizados para ajustar os parâmetros do modelo.

Por outro lado, como mostrado na figura (16) o objetivo do módulo das redes neurais artificiais é aprender uma “curva contínua” que representa os “padrões” das infinitas vistas do objeto, e simultaneamente aprender determinados agrupamentos que representam uma vista do objeto em particular. Portanto, temos um problema de classificação geral (reconhecer o padrão do objeto de forma genérica), e simultaneamente um problema de classificação local (reconhecer o padrão de uma vista em particular).

Uma forma de encaixar estes problemas na metodologia geral das redes neurais é primeiramente selecionando as funções de ativação dos neurônios da

saída da rede neural apropriados para este problema e na seqüência identificar o conjunto de exemplos de treinamento para o ajuste dos parâmetros da rede neural.

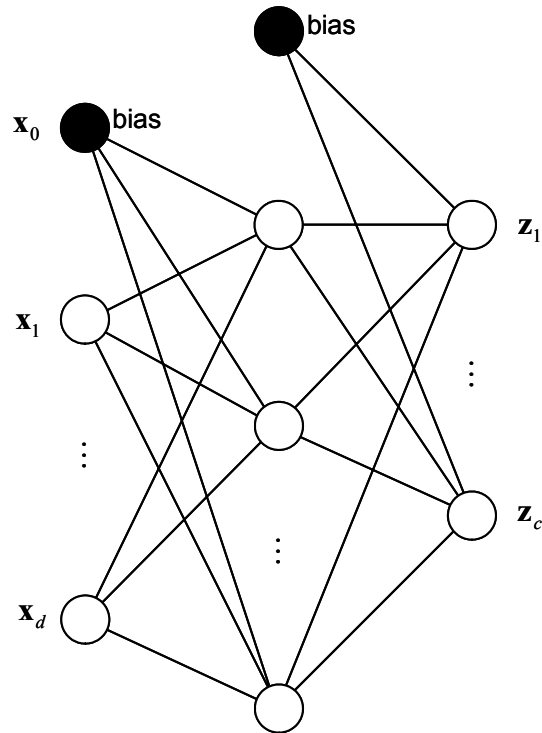


Figura 24. Arquitetura Perceptron com Duas Camadas.

A seleção da função de ativação dos neurônios da camada de saída da rede neural é importante para manter uma coerência com o tipo de problema que a rede neural pretende resolver. Para problemas de classificação a função de ativação por excelência é a função logística definida por:

$$g(a) = \frac{1}{1 + \exp(-a)} \quad (43)$$

Esta função é escolhida pelo fato de poder aproximar uma função de distribuição acumulada (figura 25).

Assim, para uma seqüência de padrões de entrada na rede neural $\mathbf{x}_{1:t} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k, \dots, \mathbf{x}_t\}$; $\mathbf{x}_k \in \mathfrak{R}^d$ é possível associar uma seqüência de probabilidades $\mathbf{z}_{1:t} = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_k, \dots, \mathbf{z}_t\}$; $\mathbf{z}_k \in \mathfrak{R}^c$ de acordo com as vistas de interesse do objeto. Portanto, os dados de treinamento podem ser definidos pelo conjunto de pares $\mathbf{D} = \{(\mathbf{x}_1, \mathbf{z}_1); (\mathbf{x}_2, \mathbf{z}_2), \dots, (\mathbf{x}_k, \mathbf{z}_k), \dots, (\mathbf{x}_t, \mathbf{z}_t)\}$. Este conjunto de N exemplos de treinamento representa um conhecimento prévio o qual permite formular um modelo Bayesiano para a estimação dos parâmetros da rede neural.

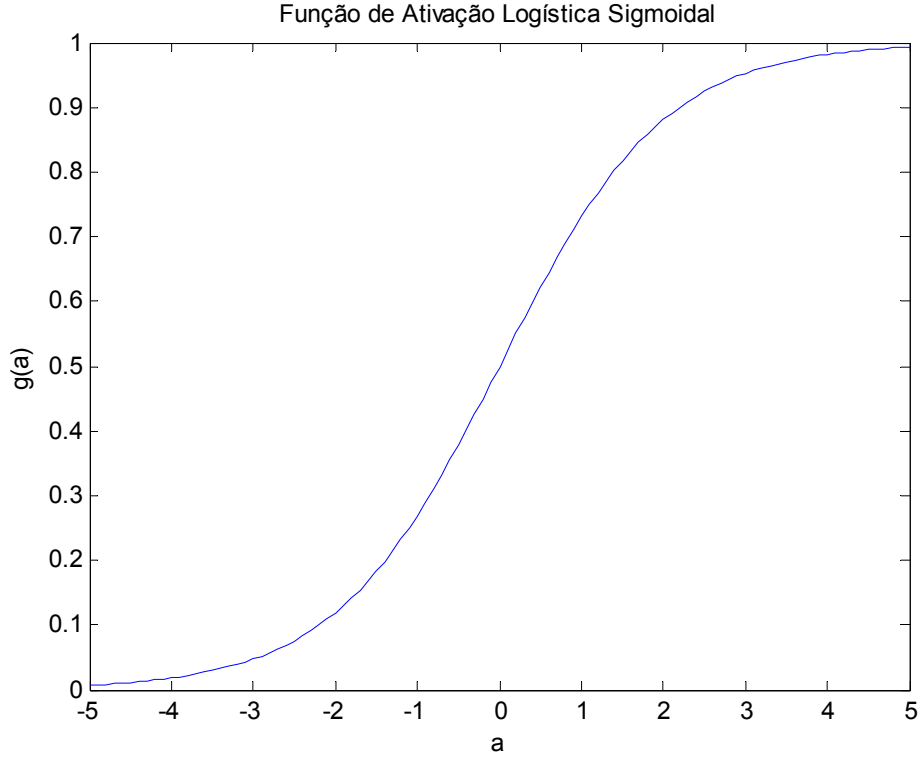


Figura 25. Função de ativação Logística Sigmoidal.

De forma genérica todos os pesos e *bias* da rede neural podem ser agrupados num vetor $\mathbf{w} \in \mathfrak{R}^W$. Assim a inferência dos parâmetros pode ser obtida através do teorema de Bayes

$$p(\mathbf{w} | \mathbf{D}) = \frac{p(\mathbf{D} | \mathbf{w})p(\mathbf{w})}{p(\mathbf{D})} \quad (44)$$

Sendo $p(\mathbf{D}) = \int p(\mathbf{D} | \mathbf{w})p(\mathbf{w})d\mathbf{w}$ um fator de normalização de forma que garante que a distribuição *a posteriori* $p(\mathbf{w} | \mathbf{D})$ possua todas as propriedades de uma distribuição de probabilidade.

A distribuição *a priori* $p(\mathbf{w})$ representa nosso conhecimento inicial sobre os valores dos parâmetros do modelo. Experimentalmente, tem se comprovado que quando o módulo parâmetros $|\mathbf{w}|$ adota valores numéricos elevados a rede apresenta problemas de generalização nas suas saídas [44], assim, a comunidade tem adotado a seguinte distribuição Gaussiana regularizada:

$$p(\mathbf{w}) = \frac{1}{\mathbf{Z}_W(\alpha)} \exp(-\alpha \mathbf{E}_W) \quad (45)$$

Sendo $\mathbf{E}_W = 0.5 |\mathbf{w}|^2$ e $\mathbf{Z}_W(\alpha) = (2\pi/\alpha)^{W/2}$.

Entretanto, na literatura são propostas outras distribuições *a priori* como por exemplo, a *a priori* Laplaciana [48] da forma $\mathbf{E}_W = \sum_i |\mathbf{w}_i|$, e a *a priori* baseada no conceito de entropia [49]. Em [50] são propostas distribuições *a priori* para redes com elevado número de neurônios.

Com relação à distribuição de verossimilhança $p(\mathbf{D} | \mathbf{w})$, pode ser adotada uma distribuição Gaussiana para avaliar a aderência do vetor de parâmetros \mathbf{w} no mapeamento dos padrões de entrada e de saída realizado pelo modelo neural. Sendo as saídas do modelo condicionalmente independentes aos padrões de entradas, então a distribuição de verossimilhança $p(\mathbf{D} | \mathbf{w}) = \prod_{k=1}^N p(\mathbf{z}_k | \mathbf{w}_k, \mathbf{x}_k)$ pode ser expressa genericamente por:

$$p(\mathbf{D} | \mathbf{w}) = \frac{1}{\mathbf{Z}_D(\beta)} \exp(-\beta \mathbf{E}_D) \quad (46)$$

Sendo $\mathbf{E}_D = 0.5 \sum_{k=1}^N (g(\mathbf{x}, \mathbf{w}) - \mathbf{z}_k)^2$; $\mathbf{Z}_D(\beta) = (2\pi / \beta)^{N/2}$ e $g(\mathbf{x}, \mathbf{w})$ é a saída da rede neural. Substituindo as equações (45) e (46) em (44) obtém-se:

$$p(\mathbf{w} | \mathbf{D}) = \frac{1}{\mathbf{Z}_E} \exp(-\mathbf{E}(\mathbf{w})) \quad (47)$$

Sendo $\mathbf{E}(\mathbf{w}) = \beta \mathbf{E}_D + \alpha \mathbf{E}_W$ e $\mathbf{Z}_E = \mathbf{Z}_W \mathbf{Z}_D$. Finalmente, os parâmetros da rede podem ser estimados como sendo média ou a moda $p(\mathbf{w} | \mathbf{D})$. Infelizmente, estes valores não podem ser estimados uma vez que $p(\mathbf{w} | \mathbf{D})$ possui uma forma analítica complexa. Porém, na literatura têm surgido duas propostas de solução para a estimativa dos parâmetros da rede neural: a primeira proposta é baseada na moda [45] a qual é estimada por um processo de otimização sobre uma aproximação Gaussiana de $p(\mathbf{w} | \mathbf{D})$; e na segunda proposta é baseada na média [50] a qual é estimada por simulações estocásticas $p(\mathbf{w} | \mathbf{D})$.

Observe-se que a distribuição $p(\mathbf{w} | \mathbf{D})$ possui um formato exponencial equivalente à equação (2). Logo os valores ótimos (média ou moda) dos parâmetros \mathbf{w} pode ser obtido através do processo de otimização ou de amostragem descrito na Seção 2.3.1. Com isto, este trabalho resulta numa terceira proposta de solução da estimativa dos parâmetros da rede neural, com a vantagem de que pode-se estimar a média ou a moda, uma vez que o filtro evolutivo é um amostrador e otimizador ao mesmo tempo.

Até aqui assumimos conhecidos os valores de α , β e W . Entretanto, estes valores, denominados comumente de hiper-parâmetros, são desconhecidos. Assim, a seguinte figura mostra todos os valores desconhecidos (parâmetros e hiper-parâmetros) do módulo de redes neurais. Na próxima seção é apresentada uma forma de estimar estes hiper-parâmetros de forma automática. Para mais detalhes de outras metodologias, veja [51].

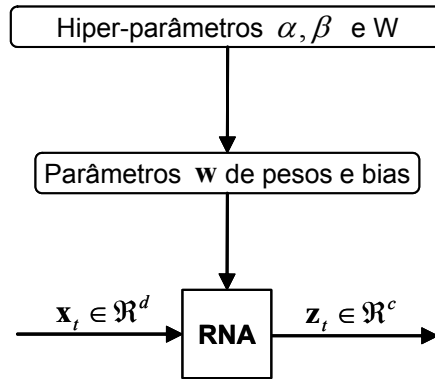


Figura 26. Parâmetros e Hiper-parâmetros do módulo de RNA.

3.2. Otimização Global do Sistema de Visão Artificial

Na seção anterior foi apresentado o modelo de visão artificial como um sistema composto de três módulos: (a) aquisição de dados, (b) pré-processamento de dados e (c) classificação de padrões (figura 16).

É um fato que a otimização dos parâmetros de cada módulo de forma isolada não garantem a otimização global do sistema. Inclusive no módulo de redes neurais os hiper-parâmetros tipicamente são determinados por tentativa e erro. A abordagem Bayesiana soluciona este problema incorporando estes hiper-parâmetros na equação da distribuição *a posteriori* dos parâmetros através do conceito de marginalização [48]-[53].

$$\begin{aligned}
 p(\mathbf{w} | \mathbf{D}) &= \iint p(\mathbf{w}, \alpha, \beta | \mathbf{D}) d\alpha d\beta & (48) \\
 &= \frac{1}{p(\mathbf{D})} \iint p(\mathbf{D} | \mathbf{w}, \alpha, \beta) p(\mathbf{w} | \alpha, \beta) p(\alpha, \beta) d\alpha d\beta \\
 &= \frac{1}{p(\mathbf{D})} \iint p(\mathbf{D} | \mathbf{w}, \beta) p(\mathbf{w} | \alpha) p(\alpha) p(\beta) d\alpha d\beta \\
 &= \frac{1}{p(\mathbf{D})} \int p(\mathbf{w} | \alpha) p(\alpha) d\alpha \int p(\mathbf{D} | \mathbf{w}, \beta) p(\beta) d\beta
 \end{aligned}$$

Note na equação anterior que a distribuição da verossimilhança não depende de α , então, $p(\mathbf{D} | \mathbf{w}, \alpha, \beta) = p(\mathbf{D} | \mathbf{w}, \beta)$. Analogamente a distribuição *a priori* não depende de β , então $p(\mathbf{w} | \alpha, \beta) = p(\mathbf{w} | \alpha)$. Além disso, na suposição que os dois hiper-parâmetros são independentes, então $p(\alpha, \beta) = p(\alpha)p(\beta)$.

Finalmente, para avaliar a integral da equação (48) é necessário realizar escolhas específicas para $p(\alpha)$ e $p(\beta)$. Sabemos que estes hiper-parâmetros estão relacionados com as variâncias da distribuição *a priori* e da distribuição da verossimilhança, conseqüentemente não podem admitir valores negativos. Por outro lado, na falta de conhecimento adicional dos valores destes hiper-parâmetros, pode-se atribuir distribuições “impróprias” ou uniformes:

$$p(\alpha) = \frac{1}{\alpha} \quad (49)$$

$$p(\beta) = \frac{1}{\beta} \quad (50)$$

Sendo α e β independentes, a equação (48) pode ser avaliada em duas partes. A integral α pode ser avaliada sobre a distribuição *a priori* $p(\mathbf{w})$ da equação (45):

$$\begin{aligned} p(\mathbf{w}) &= \int_0^{\infty} p(\mathbf{w} | \alpha) p(\alpha) d\alpha & (51) \\ &= \int_0^{\infty} \frac{1}{\mathbf{Z}_w(\alpha)} \exp(-\alpha \mathbf{E}_w) \frac{1}{\alpha} d\alpha \\ &= (2\pi)^{-\frac{W}{2}} \int_0^{\infty} \exp(-\alpha \mathbf{E}_w) \alpha^{\frac{W}{2}-1} d\alpha \\ &= \frac{\Gamma(W/2)}{(2\pi \mathbf{E}_w)^{\frac{W}{2}}} \end{aligned}$$

Sendo Γ a função gama $\Gamma(n) = \int_0^{\infty} x^{n-1} e^{-x} dx$ com $n = W/2$ e $x = \alpha \mathbf{E}_w$. De forma análoga, a integral de β pode ser avaliada sobre a distribuição da verossimilhança da equação (46), tendo como resultado:

$$p(\mathbf{D} | \mathbf{w}) = \frac{\Gamma(N/2)}{(2\pi \mathbf{E}_D)^{\frac{N}{2}}} \quad (52)$$

Substituindo os resultados da equação (51) e (52) em (48) obtém-se:

$$p(\mathbf{w} | \mathbf{D}) = \frac{1}{p(\mathbf{D})} \frac{\Gamma(N/2)}{(2\pi\mathbf{E}_D)^{\frac{N}{2}}} \frac{\Gamma(W/2)}{(2\pi\mathbf{E}_W)^{\frac{W}{2}}} \quad (53)$$

Contrastando a equação anterior com a equação (2), a função de custo a ser minimizado na equação é:

$$-\ln p(\mathbf{w} | \mathbf{D}) = \frac{N}{2} \ln \mathbf{E}_D + \frac{W}{2} \ln \mathbf{E}_W + \text{constante} \quad (54)$$

Todavia o gradiente desta função de custo é dado por:

$$-\nabla \ln p(\mathbf{w} | \mathbf{D}) = \frac{N}{2\mathbf{E}_D} \nabla \mathbf{E}_D + \frac{W}{2\mathbf{E}_W} \nabla \mathbf{E}_W \quad (55)$$

Analogamente, a função de custo a ser minimizado na equação (47), assim como o seu gradiente são expressos pela equações respectivamente

$$-\ln p(\mathbf{w} | \mathbf{D}) = \beta \mathbf{E}_D + \alpha \mathbf{E}_W + \text{constante} \quad (56)$$

$$-\nabla \ln p(\mathbf{w} | \mathbf{D}) = \beta \nabla \mathbf{E}_D + \alpha \nabla \mathbf{E}_W \quad (57)$$

Finalmente, contrastando as equações (55) e (57) obtemos

$$\alpha_{\text{ótimo}} = \frac{W}{2\mathbf{E}_W} \quad (58)$$

$$\beta_{\text{ótimo}} = \frac{N}{2\mathbf{E}_D} \quad (59)$$

Com este resultado, comprova-se que a metodologia Bayesiana, permite realizar a inferência completa dos parâmetros e hiper-parâmetros de qualquer modelo embora neste caso em particular tenha-se tratado o caso do modelo neural.

Note-se que a solução anterior poderia ser obtido numericamente como a otimização da função de custo da equação (54)² ou da otimização da função de custo da equação (56). Generalizando este raciocínio de cálculo de hiper-parâmetros, todos os parâmetros do módulo SMO e ACP podem ser vistos como hiper-parâmetros para o módulo de RNA. Como no caso anterior, temos duas forma de estimar estes hiper-parâmetros do modelo de visão computacional. A primeira forma é através da otimização da função de custo resultante da reformulação da equação (48)

$$p(\mathbf{w} | \mathbf{D}) = \int p(\mathbf{w}, \varphi | \mathbf{D}) d\varphi \quad (60)$$

Sendo $\varphi = (\text{parâmetros RNA}, \text{parâmetros SMO}, \text{parâmetros ACP})$.

Aplicando o teorema de Bayes na distribuição conjunta $p(\mathbf{w}, \varphi | \mathbf{D})$ temos:

$$p(\mathbf{w} | \mathbf{D}) = \frac{1}{p(\mathbf{D})} \int p(\mathbf{D} | \mathbf{w}, \varphi) p(\mathbf{w}, \varphi) d\varphi \quad (61)$$

Expandindo a distribuição conjunta $p(\mathbf{w}, \varphi)$ obtém-se:

$$p(\mathbf{w} | \mathbf{D}) = \frac{1}{p(\mathbf{D})} \int p(\mathbf{D} | \mathbf{w}, \varphi) p(\mathbf{w} | \varphi) p(\varphi) d\varphi \quad (62)$$

Ainda, a integral da equação anterior, pode ser aproximado através de simulações Monte Carlo baseado num conjunto finito de N amostras $\{\varphi^{(i)}, i = 1, \dots, N\}$ da distribuição $p(\varphi)$. Logo a função de custo a ser otimizado (maximizado) seria :

$$p(\mathbf{w} | \mathbf{D}) \cong \frac{1}{p(\mathbf{D})} \sum_{i=1}^N p(\mathbf{D} | \mathbf{w}, \varphi^{(i)}) p(\mathbf{w} | \varphi^{(i)}) \quad (63)$$

Evidentemente que esta primeira forma de solução é complicada. Porém a segunda forma de cálculo de hiper-parâmetros seria a mais simples. Neste segundo caso a função de custo a ser otimizado (maximizado) é a função de custo da equação (47) que equivaleria à minimização da função de custo da equação (57). Note-se que em ambas funções de custo da equação (47) e (57) não aparecem explicitamente em função dos parâmetros dos módulos SMO nem ACP, mas eles possuem uma dependência indireta, uma vez que o conjunto \mathbf{D} é formado com a participação destes módulos. Conclui-se, por tanto que os valores ótimos dos parâmetros e hiper-parâmetros são aqueles que maximizam a distribuição a posteriori de $p(\mathbf{w} | \mathbf{D})$ e com isto fica mostrado que a metodologia Bayesiana permite estimar de forma global e integrada todos os parâmetros e hiper-parâmetros do sistema de visão artificial o qual foi um dos objetivos propostos e alcançados neste trabalho.

Para a comprovação da metodologia exposta neste trabalho, no próximo capítulo foram desenvolvidas duas aplicações de reconhecimento de objetos.

² Note que esta função de custo depende de forma “Indireta” dos valores de α e β

4 Aplicações do Modelo de Visão Artificial

Este capítulo tem por objetivo testar a eficiência do modelo de visão artificial proposto neste trabalho. Na primeira seção, apresenta-se a aplicação do modelo no reconhecimento de um cone tridimensional. E na segunda seção, apresenta-se a aplicação de modelo num caso real de reconhecimento de uma “cabeça de poço” que é um equipamento utilizado na produção de petróleo numa jazida localizada no mar.

4.1. Aplicação de Visão de Objetos Tridimensionais

Nesta primeira aplicação, foi desenvolvido um aplicativo em C++, como mostrado na seguinte figura, que permite através de controles (situados à direita da janela da figura 27) renderizar diversas vistas de um cone tridimensional.

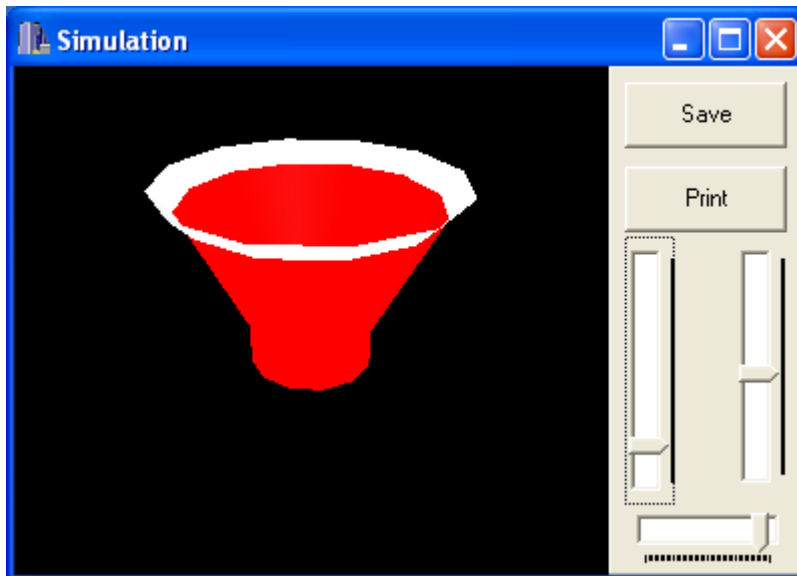


Figura 27. Aplicativo de renderização de um cone tridimensional.

Sendo o cone uma figura simétrica, então só é possível ter vistas diferentes para ângulos de variação entre de 0° a 90° (as vistas dos ângulos entre 90° a 180° são equivalentes às vistas dos ângulos entre 0° a 90°).

Logo, a faixa de vistas entre 0° a 90° foram divididos em 6 faixas: $[0^\circ\sim 15^\circ[$; $[15^\circ\sim 30^\circ[$; $[30^\circ\sim 45^\circ[$; $[45^\circ\sim 60^\circ[$; $[60^\circ\sim 75^\circ[$; $[75^\circ\sim 90^\circ[$. Por cada faixa foram extraídas 5 vistas de forma aleatória. A seguinte seqüência de figuras mostra as 6 faixas de vistas cada qual contendo 5 imagens de tamanho (64 x 64) pixels.

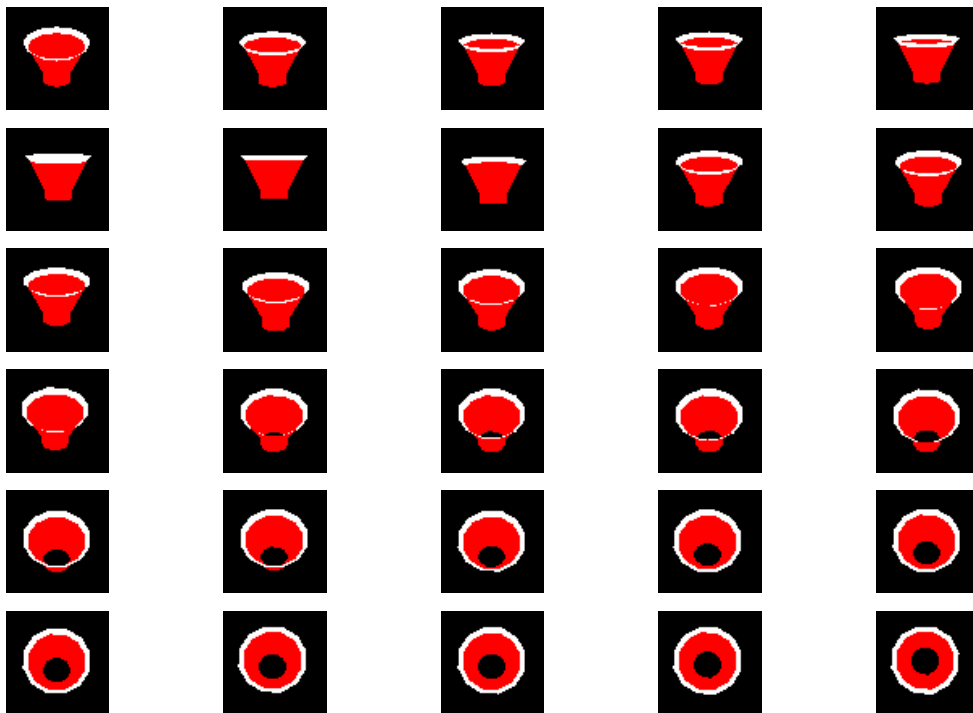


Figura 28. Vistas do cone tridimensional.

A idéia principal deste teste é anular o módulo de aquisição de dados, gerando imagens de forma artificial, para concentrar-se exclusivamente nos módulos de pre-processamento e no módulo de classificação de padrões. Desta forma, avaliar melhor o desempenho do modelo de visão artificial.

Aplicando o algoritmo descrito no capítulo anterior, foi otimizado o sistema de forma simultaneamente tendo como resultado: uma arquitetura de rede neural de 10 entradas (10 componentes principais) ;16 neurônios na camada escondida e 6 neurônios na camada de saída (um para cada agrupamento de vistas). Além disso foi definido como probabilidade de sucesso (na saída da rede neural) valores acima de 0,9.

Após este treinamento do sistema, posteriormente foi introduzido 100 imagens escolhidas de forma aleatória do banco de dados. Como resultado o sistema teve 100% de acerto no reconhecimento do grupo (vista do objeto) da qual a imagem era parte.

4.2. Aplicação de Visão Submarina na Indústria de Petróleo

Atualmente o petróleo tem-se tornado uma das fontes energéticas mais importantes no planeta. A demanda crescente deste recurso tem incentivado as diversas companhias de petróleo na exploração em novas fronteiras, geralmente no mar e em lâminas d'água cada vez mais profundas, como mostrado na seguinte figura (recordes de prospecção da Petrobrás).

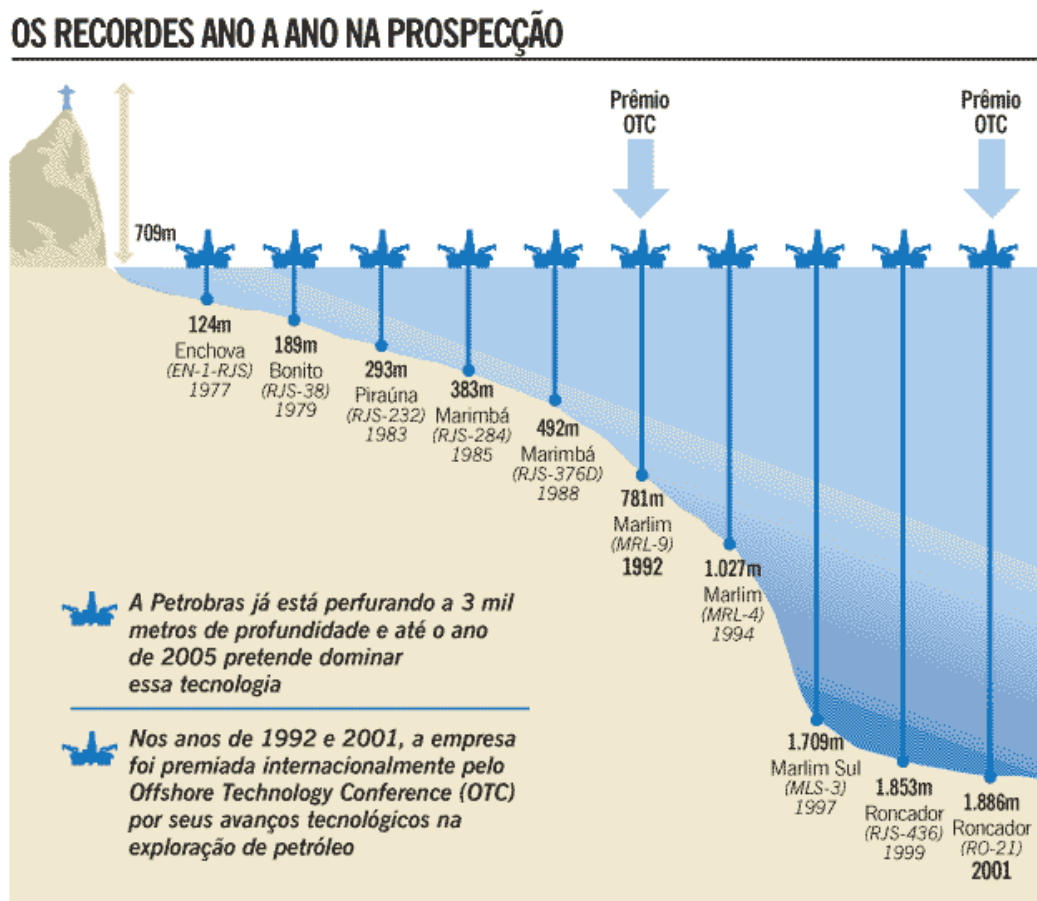


Figura 29. Recordes da Petrobras na exploração offshore.

Qualquer operação de exploração ou produção de petróleo em profundidades acima de 300 metros deve ser realizada utilizando sistemas que não

dependam da intervenção humana de mergulhadores. Assim, essas operações devem necessariamente ser totalmente automatizadas com o auxílio de robôs submarinos, conhecidos como veículos operados remotamente (VOR's). A figura (30) mostra a quantidade de operações de instalação e desinstalação de equipamentos envolvidos na produção de um campo de petróleo.

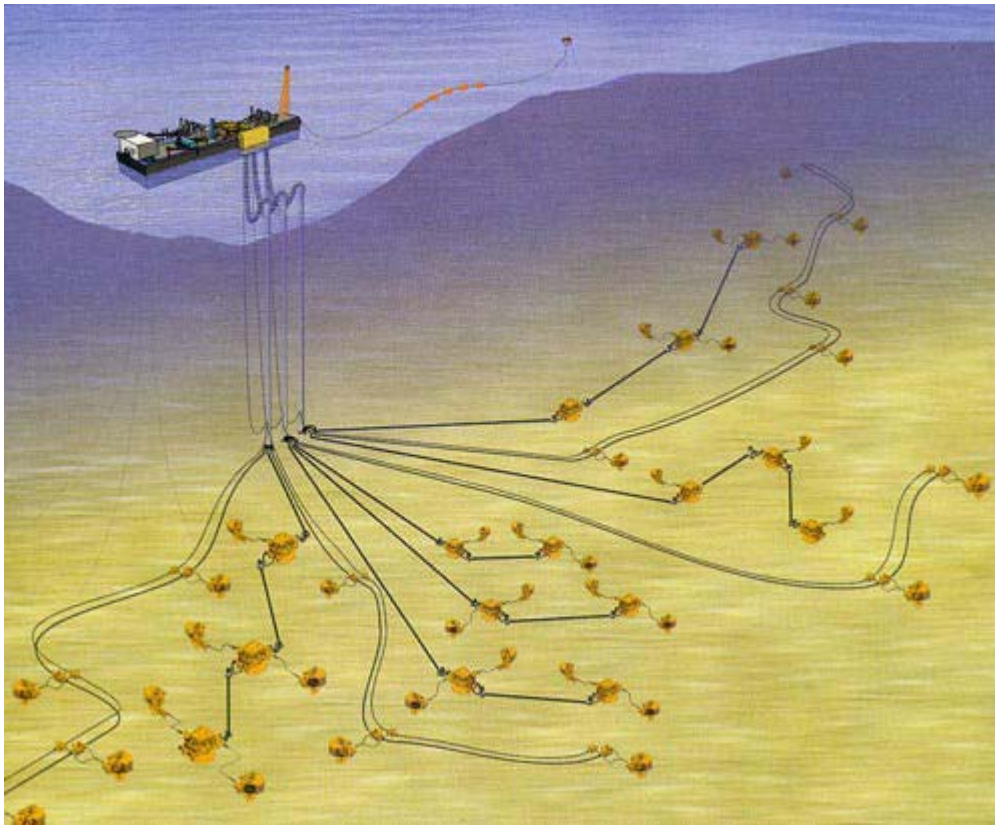


Figura 30. Equipamentos num campo de produção de petróleo.

Evidentemente, mesmo com o auxílio dos VOR's as operações de instalação e desinstalação de equipamentos no solo submarino tornam-se demoradas, elevando o custo da exploração ou produção de petróleo. Para se ter uma noção destes custos, o aluguel de uma sonda (plataforma) de perfuração custa atualmente de 200 mil a 400 mil dólares por dia. Portanto, qualquer otimização neste tempo resulta numa economia de tempo e recursos para a companhia de petróleo.

Neste contexto, um sistema de visão artificial com a capacidade de reconhecimento de equipamentos pode representar numa ferramenta aceleradora das operações, contribuindo desta forma na redução do tempo das operações e como consequência redução dos custos operacionais.

Entre os equipamentos mais operados estão as “árvores de natal” e as “cabeças de poços” [57]. Neste trabalho, foi escolhido o reconhecimento do equipamento da cabeça de poço, a qual é mostrada na parte central da figura (31). Esta figura provém da câmera de um robô operando no fundo do mar. Na parte inferior direita pode ser visto o braço mecânico do robô.



Figura 31. Equipamento de cabeça de poço.

A seqüência de imagens geradas nestas condições possui muito ruído em decorrência da suspensão de areia do fundo do mar, iluminação dinâmica, etc. Portanto, surge a necessidade de filtrar estas imagens com o objetivo de uniformizá-las. A figura (32) mostra a imagem anterior processada com filtros de remoção de ruído e aumento de contraste.



Figura 32. Imagem filtrada.

Após esta filtragem é necessário segmentar os objetos contidos na imagem. Este isolamento pode ser feito, inicialmente, executando um algoritmo de filtragem de “identificação de bordas”, como mostrado na figura (33).



Figura 33. Detecção de borda dos objetos.

Seguir-se-á de um algoritmo de clusterização. Após este isolamento de objetos, volta-se à imagem filtrada para extrair a região correspondente ao objeto de interesse com mostrado na figura (34).

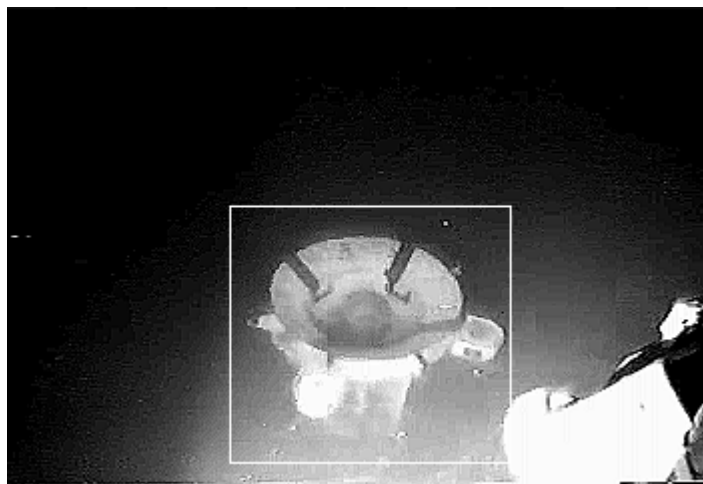


Figura 34. Isolamento do objeto de interesse.

Logo, procede-se a padronizar a imagem da região de interesse para um tamanho de (64 X 64) pixels. A seguinte seqüência de figuras mostra as imagens padronizadas, para cinco agrupamentos de vistas (cada qual composto de 60 imagens), todos eles prontos para entrar no módulo de PCA e RNA.

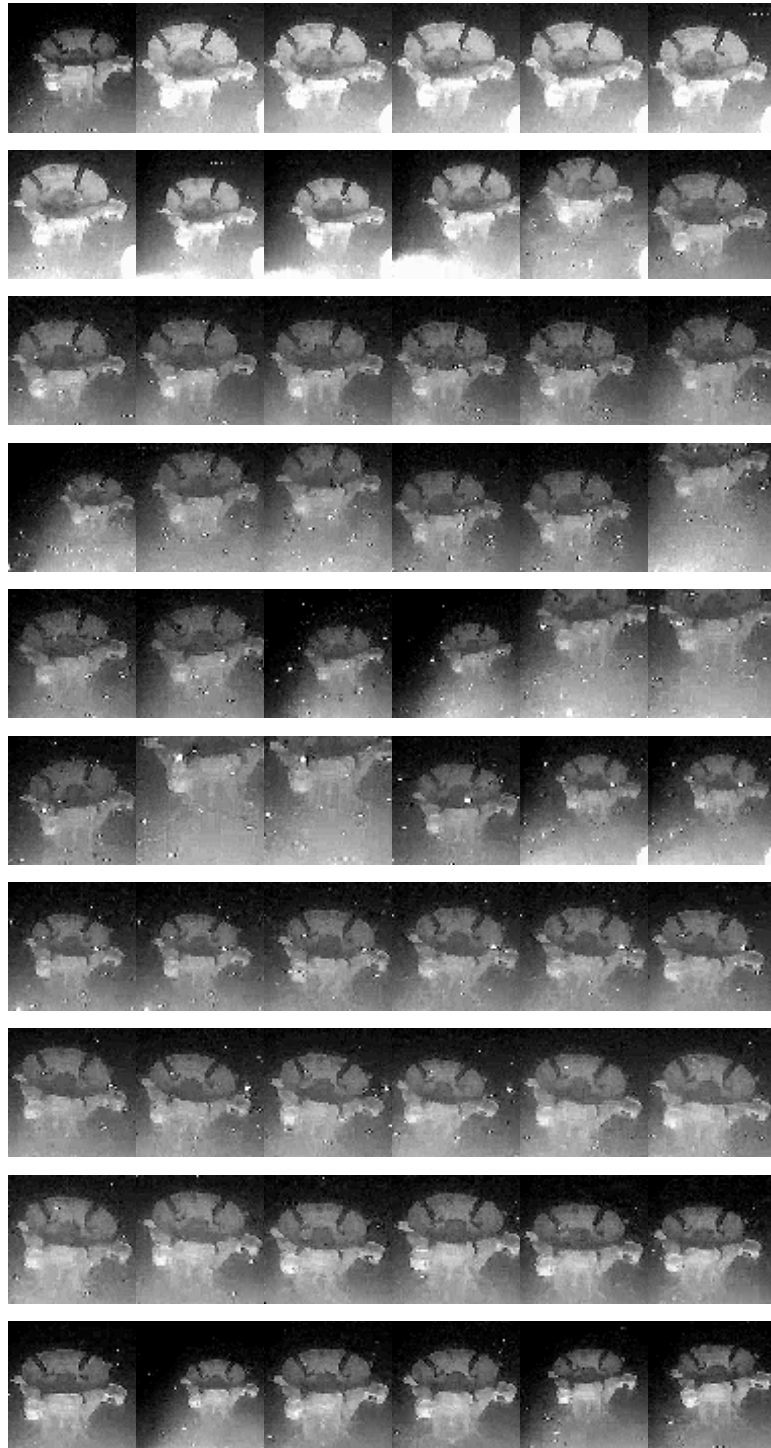


Figura 35. Padrões da primeira vista do objeto.

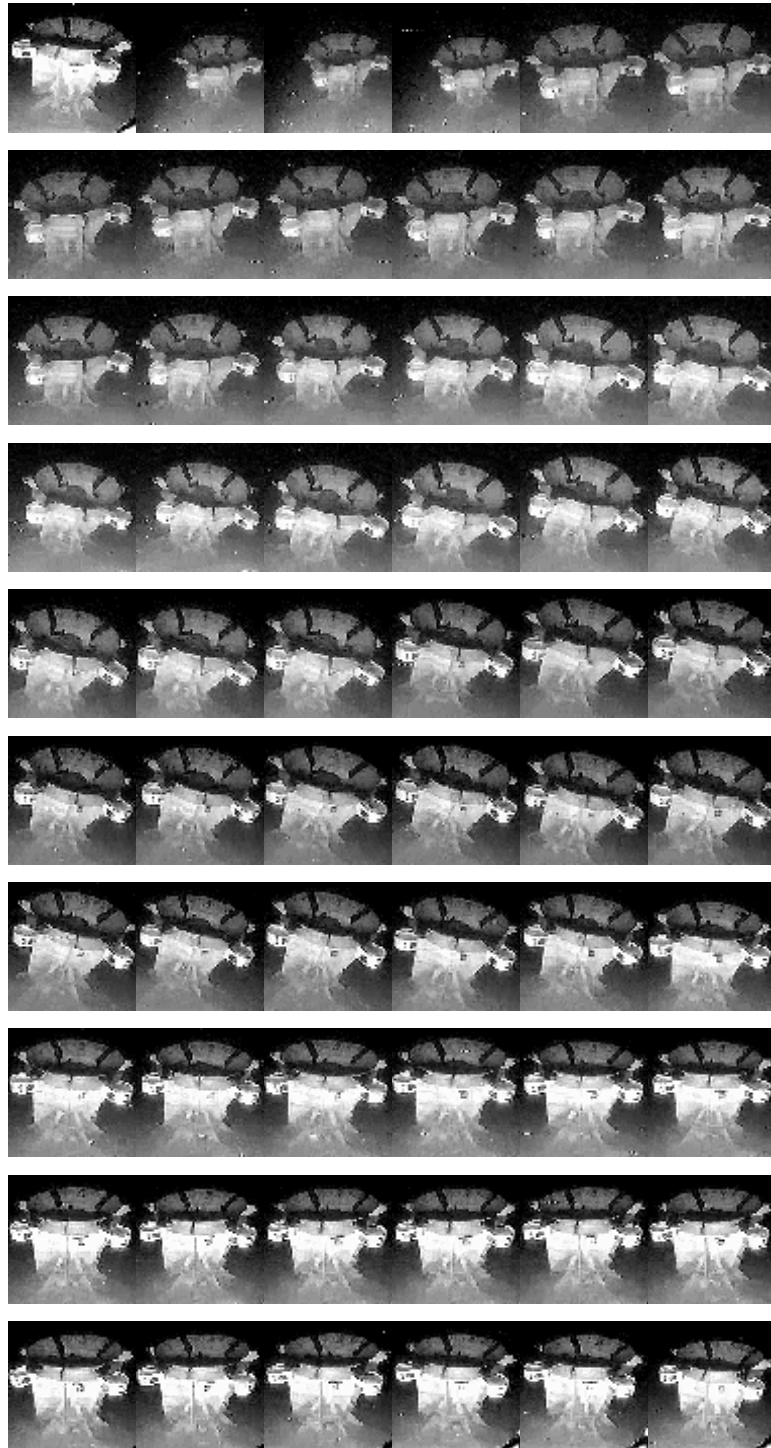


Figura 36. Padrões da segunda vista do objeto.

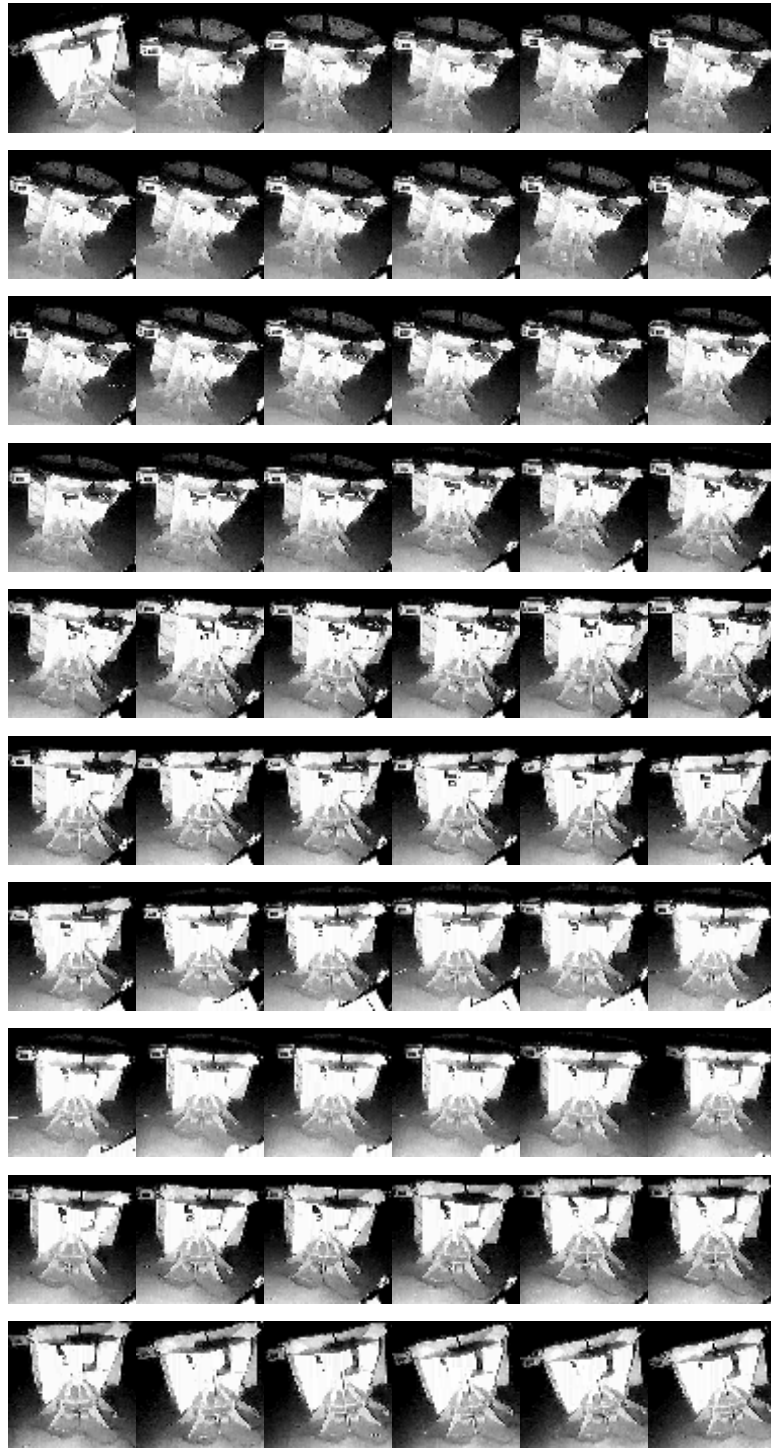


Figura 37. Padrões da terceira vista do objeto.

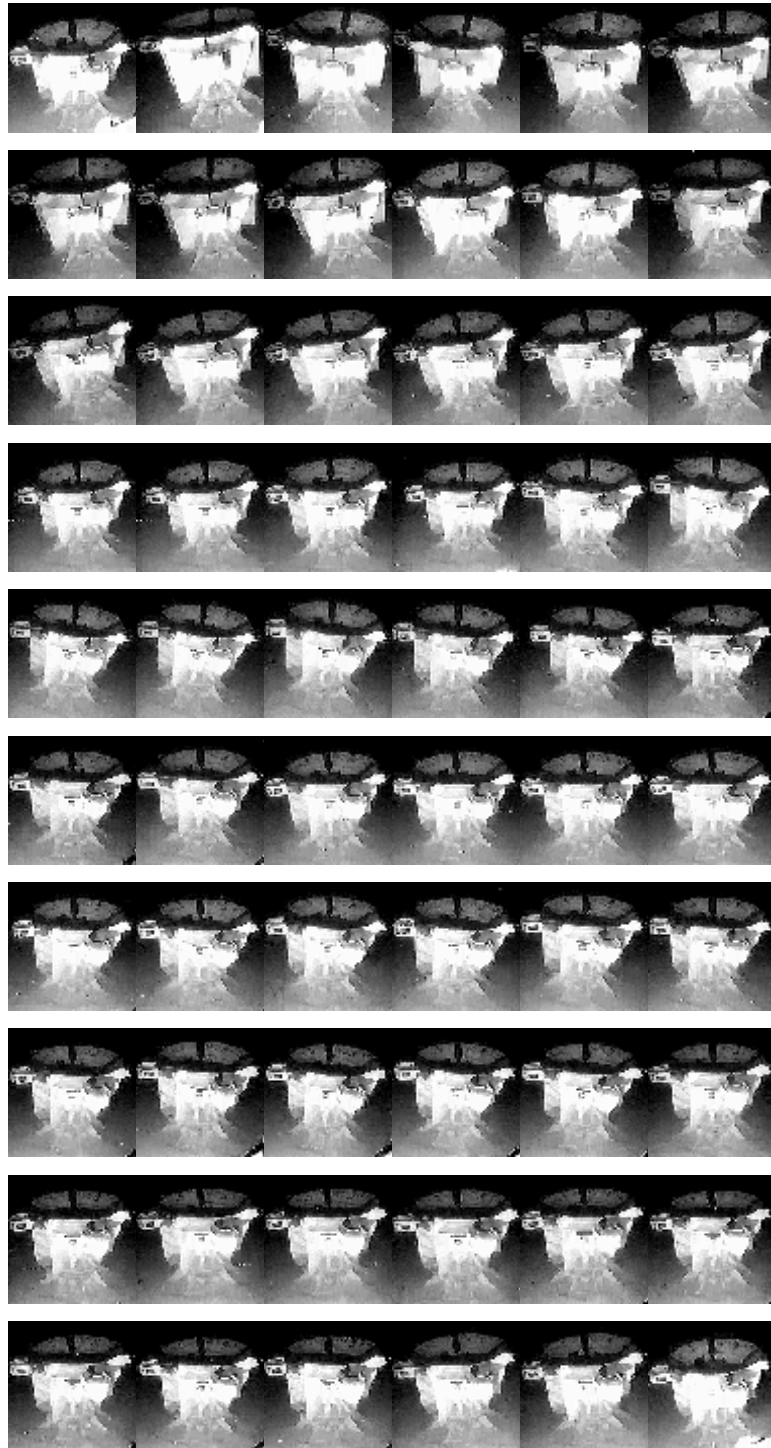


Figura 38. Padrões da quarta vista do objeto.



Figura 39. Padrões da quinta vista do objeto.

Aplicando o algoritmo descrito no capítulo anterior, foi otimizado o sistema de visão artificial de forma simultânea, tendo como resultado: uma arquitetura de rede neural de 8 entradas (8 componentes principais); 13 neurônios na camada escondida e 5 neurônios na camada de saída (um para cada agrupamento de vistas). Além disso, foram definidos como probabilidade de sucesso (na saída da rede neural) valores acima de 0,9.

Para testar a eficiência do sistema foram introduzidas 1000 imagens escolhidas de forma aleatória das imagens do banco de dados. O sistema teve como resultado 82% de acerto no reconhecimento do grupo (vista do objeto) da qual a imagem era parte e 97,5% no reconhecimento do objeto como um todo.

Por ultimo, foi realizado um teste do sistema no modo *on-line*, ou seja a entrada do sistema foi acoplado diretamente no vídeo proveniente do objeto. Para visualizar os resultado deste, toda vez que o sistema detectava algum padrão do objeto, o vídeo original era alterado para adicionar na imagem um quadrado em torno da região onde o objeto foi reconhecido. É importante destacar que para acelerar o processamento no modo de operação *online* é necessário ter um sub-módulo de seguimento ou *tracking* do objeto (como mencionado na Seção 2.3.3) que tem como objetivo “focalizar” a atenção do sistema em torno de uma região da imagem, pois do contrario, o sistema teria sempre que processar “toda” a imagem. Neste trabalho foi incorporado um sub-módulo simples de seguimento que consistiu basicamente no aumento da variância em torno do padrão reconhecido. Na pratica isto equivaleria a “focalizar” a atenção do sistema em torno da região atual do objeto. As seguintes figuras mostram algumas imagens do vídeo neste teste *on-line*.

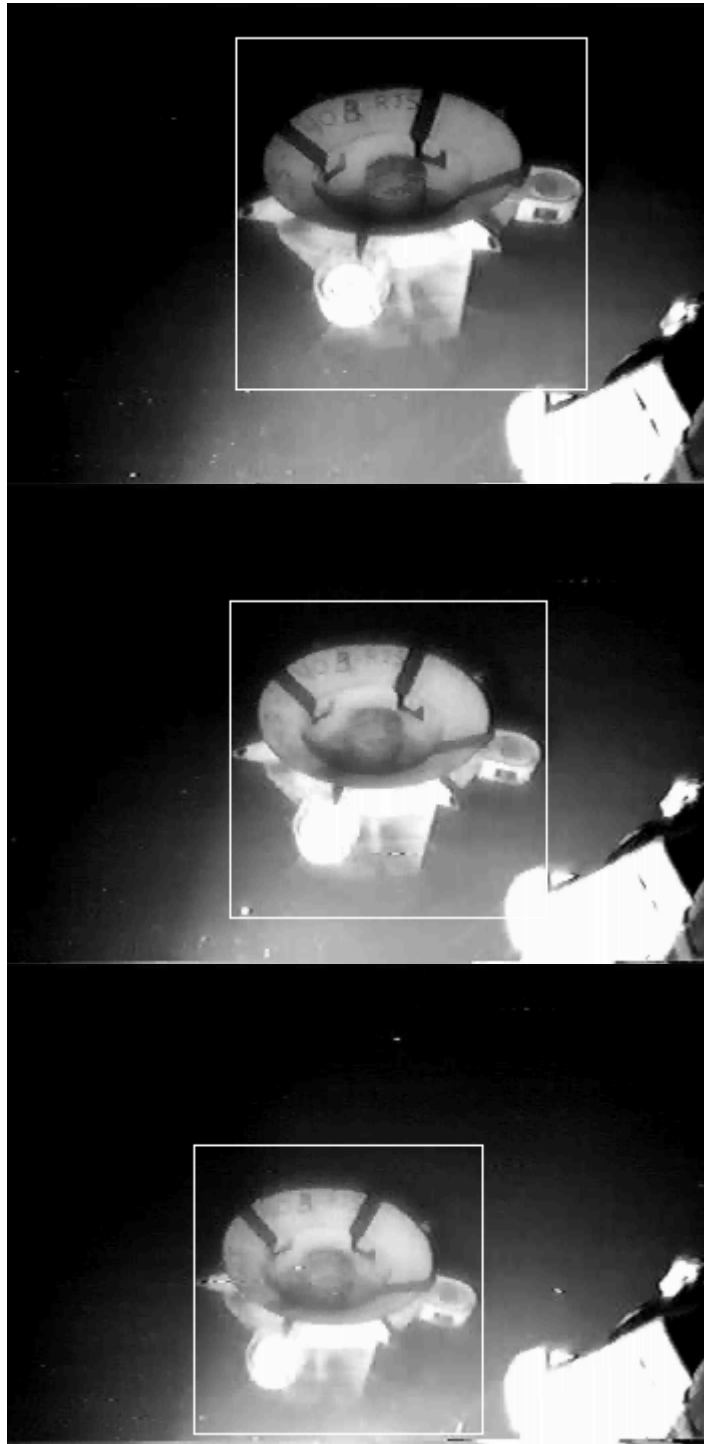


Figura 40. Reconhecimento *on-line* do objeto.

4.3.

Detalhes da Implementação Computacional do Algoritmo

A seguir especificam-se os detalhes computacionais do algoritmo utilizado na aplicação descrita na Seção 4.2. O mesmo algoritmo vale para a aplicação descrita na Seção 4.1, com a ligeira exceção de que esta aplicação não possui módulo SMO.

Evidentemente, existem algumas peculiaridades próprias ou específicas destas aplicações, mas o espírito da mecânica de estimação Bayesiana foi preservada. Além disso, alguns detalhes são inerentes à plataforma de implementação. Neste caso, foram utilizados o ambiente e rotinas disponíveis do Matlab.

Começaremos com os detalhes do Sistema Mono-Ocular (SMO). Não foi considerado nenhum parâmetro relacionado com o sensor ótico. Também não foi considerado os parâmetros referente ao filtro de “aumento de contraste”. Já o filtro de “detecção da borda” [54],[55] foi identificado como um filtro crítico devido a pode detectar um conjunto de gradientes espúrios dos ruídos. Para contornar este problema foi adicionado um filtro morfológico que tem o objetivo de eliminar grupos de pixels que possuem uma área pequena e que poderia ser considerada como ruído. Entretanto o valor limite ℓ que a partir os pixels são filtrados torna-se um parâmetro que precisa ser otimizado.

Os detalhes da análise de componentes estão especificados na Seção 3.1.2, tendo como único parâmetro o número de componentes principais η .

Finalmente, os detalhes do módulo de redes neurais estão especificadas na Seção 3.1.3 tendo como hiper-parâmetros α , β e W , e como parâmetros os pesos e *bias* \mathbf{w} .

Note que os parâmetros dos módulos SMO e ACP podem ser vistos como hiper-parâmetros da rede neural como vistos na Seção 3.2.

A figura (41) mostra os detalhes dos parâmetros e hiper-parâmetros do sistema de visão computacional utilizados nas aplicações descritas na seção anterior. Note-se que existe uma dependência hierárquica entre os hiper-parâmetros e parâmetros. Pos exemplo, o hiper-parâmetro W define o número de parâmetros da rede neural.

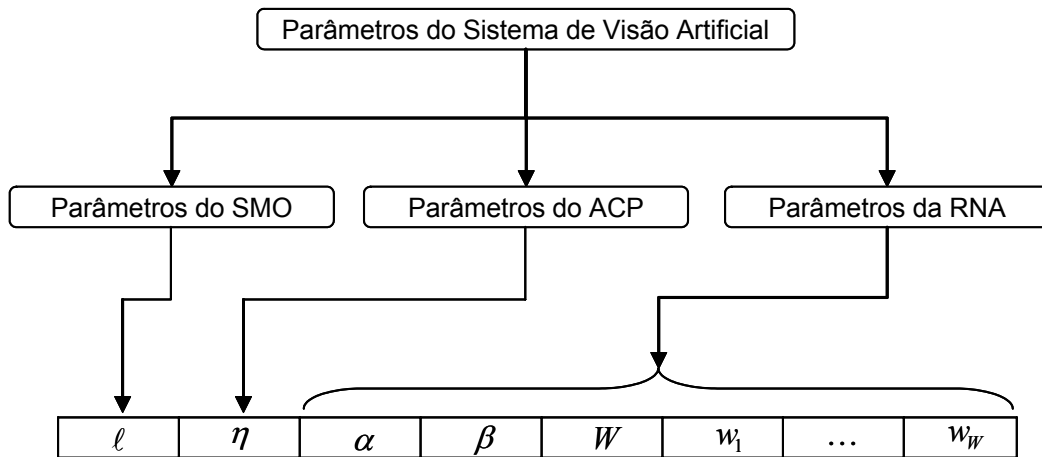


Figura 41. Detalhes dos Parâmetros do Sistema de Visão Artificial.

Por esta razão, a otimização global do sistema na prática é resolvida através de dois otimizadores, como mostrado na figura (42). A princípio poderia ser utilizado o filtro evolutivo para os dois otimizadores. Entretanto, esta combinação seria pesada computacionalmente, de forma que na prática o segundo otimizador pode ser, por exemplo, um otimizador gradiente escalado conjugado.

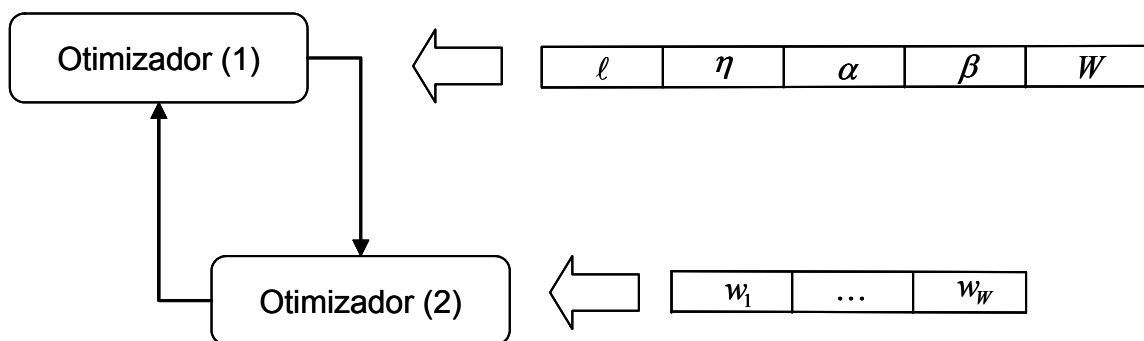


Figura 42. Otimização do sistema de visão artificial.

Como mostrado na Seção 3.2, a função de custo a ser maximizada pelos dois otimizadores é dada pela equação (47), sendo que durante a execução do primeiro otimizador os valores dos parâmetros assumem valores constantes, e durante a execução do segundo otimizador os valores dos hiper-parâmetros assumem valores constantes.

Com estas considerações, os detalhes do algoritmo do primeiro otimizador começam com a definição do cromossoma, o qual estaria formado pelo vetor de hiper-parâmetros $\theta = [l, \eta, \alpha, \beta, W]$ com $l \geq 0$, $\eta \geq 0$, $\alpha \geq 0$, $\beta \geq 0$, $W \geq 0$. É importante lembrar que o valor de W é um numero que depende do número de

neurônios da camada de entrada da rede neural nIn , o número de neurônios na camada escondida $nHidden$ e o número de neurônios da camada de saída $nOut$, ou seja: $W = (nIn + 1) * nHidden + (nHidden + 1) * nOut$. O número de neurônios da camada de entrada é igual ao número de componentes principais, ou seja, $nIn = \eta$. O número de neurônios na camada de saída é $nOut = 5$. Todavia, afim de evitar problemas numéricos, a distribuição inicial dos hiper-parâmetros $\theta_0 = [\ell_0, \eta_0, \alpha_0, \beta_0, W_0]$ foi restringido de acordo com: $0 \leq \ell_0 \leq 1$, $0 \leq \eta_0 \leq 20$, $0 \leq \alpha_0 \leq 1$, $0 \leq \beta_0 \leq 1$, $0 \leq W_0 \leq 525$.

Logo, considerando uma população de 100 indivíduos, os estágios “Iniciar” e “Avaliar” do filtro evolutivo consistem em:

Iniciar $\{\theta_t^{*(i)}\}_{i=1}^N$:

Para $i = 1, \dots, N$, cromossomas $\theta_1^{*(i)} = \theta_0^{(i)}$

sendo $\theta_0^{(i)} \in .[\ell_0, \eta_0, \alpha_0, \beta_0, W_0]$

Avaliar $\{\theta_t^{*(i)}\}_{i=1}^N$:

Para $i = 1, \dots, N$, cromossomas

- (a) Executar o módulo SMO com o valor de $\ell^{(i)}$ correspondente ao cromossoma i .
- (b) Executar o módulo ACP com o valor de $\eta^{(i)}$ correspondente ao cromossoma i .
- (c) Executar o módulo RNA com os valores de $\alpha^{(i)}$, $\beta^{(i)}$, $W^{(i)}$ correspondente ao cromossoma i . Note que neste ponto será lançado o segundo otimizador encarregado de calcular os parâmetros $\mathbf{w}^{(i)}$ da rede neural.
- (d) Finalmente o peso de importância normalizado é avaliado por:

$$\tilde{\omega}_t^{(i)} = \frac{p(\mathbf{w}_t^{*(i)} | \mathbf{D})}{\sum_{j=1}^N p(\mathbf{w}_t^{*(j)} | \mathbf{D})} \quad (64)$$

O estágio de “Seleção” permanece inalterado. Já o estágio de “Evolução” precisa incluir um processo de verificação dos limites do domínio de cada solução. Caso exista alguma variável que ultrapasse estes limites, então esta variável é reiniciada dentro do domínio correspondente.

Na implementação do segundo otimizador, executado dentro do estágio “Avaliar” do primeiro otimizador no caso das aplicações anteriores, foi utilizado um otimizador gradiente escalado conjugado [54]. Entretanto, pode perfeitamente ser utilizado um outro filtro evolutivo, só que esta configuração é computacionalmente mais complexa.

5 Conclusões e Trabalhos Futuros

Nos últimos anos, o avanço tecnológico, tem disponibilizado ao mercado uma enorme capacidade computacional a preços acessíveis. Todo este poder computacional tem permitido implementar diversos sistemas de visão computacional visando à substituição do ser humano em tarefas consideradas perigosas, mecânicas e/ou repetitivas.

Embora a concepção do problema de visão artificial esteja claramente definida, ou seja, imitar as habilidades da visão humana, a procura pela sua solução se mostrou um enorme desafio. Diversas áreas da ciência já têm contribuído com estudos focalizados sobre este problema. Entretanto, a área de *Inteligência Artificial* é a que tem abraçado a missão da procura da solução deste problema. Neste contexto, este trabalho pretende contribuir nesta área ao propor uma metodologia que visa integrar o processo de estimação dos parâmetros do modelo de visão artificial.

Entretanto, a escolha do modelo paramétrico foi baseada nos conhecimentos acumulados pelas diversas disciplinas e teorias da ciência relacionadas com o tema. Assim, foi escolhido o modelo de redes neurais (arquitetura percentron com múltiplas camadas) como uma forma de imitar os neurônios do cérebro humano. Além disso, como sub-módulos auxiliares de pré-processamento e aquisição de dados, foram escolhidas metodologias consagradas na literatura, como a Análise de Componentes Principais (ACP) e a metodologia de detecção de bordas.

Com base neste modelo de visão artificial, procedeu-se a aplicação de um algoritmo de otimização global baseado na teoria Bayesiana e na teoria de evolução de Darwin, como mostrado no Capítulo 2.

Na comprovação do modelo proposto, foram desenvolvidas duas aplicações: a primeira trata do reconhecimento de objetos tridimensionais e a segunda de uma aplicação real de visão submarina. Com base nos resultados apresentados, podemos afirmar que o sistema de visão artificial mostra-se promissor e viável para uma aplicação comercial.

A contribuição em maior destaque neste trabalho se refere ao relacionamento da equivalência entre o princípio Darwiniano da “sobrevivência dos mais aptos” e o teorema de Bayes. A partir desta equivalência foi possível adaptar determinados métodos de simulações estocásticas numa visão Darwiniana (adicionado da genética Mendeliana). Como resultado, obtêm-se algoritmos de otimização e/ou algoritmos de amostragem mais flexíveis, e capazes de resolver problemas complexos de engenharia.

Como trabalhos de extensão propõem-se:

- Generalizar o módulo de aquisição de dados de maneira que não seja muito dependente da aplicação. Por exemplo, o algoritmo de clusterização de objetos.
- Substituir o módulo de PCA convencional por outros métodos de redução de dimensionalidade. Por exemplo, poderia ser utilizada outra rede neural.
- Substituir o módulo de classificação de padrões, realizado pelas redes neurais, por outros métodos, como por exemplo, a análise de discriminantes.
- Desenvolver novas aplicações do modelo de visão artificial proposto neste trabalho.
- Tendo em vista que cada módulo do sistema de visão computacional possui diversos candidatos, propõe-se incluir esta diversidade de métodos no algoritmo de otimização global, de maneira que ele evolua para o método apropriado para cada módulo, além de encontrar os parâmetros e hiper-parâmetros do sistema.

6

Referências Bibliográficas

- [1] JAIN A.K.; DUIN R.P.W.; MAO J. Statistical Pattern Recognition: A Review; IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol.22 NO 1, January 2000.
- [2] DUDA R.O., HART P.E., STORK D.G. Pattern Classification, John Wiley & Sons Inc. 2001.
- [3] DEVIJVER P.A., KITTLER J. Pattern Classification: A Statistical Approach. Prentice Hall, Englewoods Cliffs, N.J. 1982.
- [4] FUKUNAGA K. Introduction to Statistical Pattern Recognition, second edition. Boston: Academic Press, 1990.
- [5] JAMES M. Classification Algoritms. London: Willian Collings Sons & Co. Ltda, 1995.
- [6] THOMAZ C.E. Maximum Entropy Covariance Estimate for Statistical Pattern Recognition. PhD thesis, Imperial College, 2004.
- [7] LEE T.W. Independent Component Analysis: Theory and Application, Cluber Academic Publishers. 1998.
- [8] ANDERSON T.W. An Introduction to Multivariate Statistical Analysis, second edition. New York: John Wiley & Sons, 1984.
- [9] HAYKIN S. Adaptive Filter Theory, Third Edition, Prentice Hall, 1996.
- [10] MITRA S.K. Digital Signal Processing: A computer Based Approach. Mc Graw Hill, 1998.
- [11] MARR, D. Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. W.H. Freeman and Company, 1982.
- [12] WANDELL B.A. Foundations of Vision, Sinauer Press, Sunderland, MA, 1995.
- [13] PALMER S.E. Vision Science: Photons to Phenomenology, The MIT Press, 1999.
- [14] DAWKINS, R. The Selfish Gene, Oxford University Press, New York, 1976.
- [15] PINKER S. How the Mind Works, Editora Schwardecz Ltda. 1997.
- [16] DARWIN, C.; SURIANO, G. (Editor). The Origin of Species. Grammercy; 1998.
- [17] HAWKINGS, S. Universe in a Nutshell. Bantam Doubleday Dell Pub (Trd); 1st edition; 2001.

- [18] PAPOULIS, A. Probability, Random Variables, and Stochastic Processes. Third Edition; McGRAW-HILL; 1991.
- [19] HOLLAND, J. Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbor, 1975.
- [20] MICHALEWICZ, Z. Genetic Algorithm + Data Structure = Evolution Programs. Third, Revised And Extended Edition, Springer, 1996.
- [21] GOLDBERG, D.E. Genetic Algorithm in Search, Optimization and Machine Learning. Addison-Wesley, 1989.
- [22] LIU J.S. Blind deconvolution via Sequential Imputations, Journal of the American Statistical Association, 90(430):567-576, 1995.
- [23] LIU, J.S. Monte Carlo Strategies in Scientific Computing, Springer Verlag; 2001.
- [24] LIU, J.S. Metropolized Independent Sampling with Comparison to Rejection Sampling and Importance Resampling. Statistic and Computing, 6: 113-119, 1996.
- [25] LIU, J.S.; CHEN, R. Sequential Monte Carlo Methods for Dynamic Systems. Journal of the American Statistical Association, 93: 1032-1044, 1998.
- [26] LIU, J.S. Markov Chain Monte Carlo and Related Topics. Technical Report, Department of Statistics, Stanford University, 1999.
- [27] GAMERMAN, D. Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference. Chapman & Hall, 1997.
- [28] METROPOLIS, N.; ROSENBLUTH, A.W.; ROSENBLUTH, N.W.; TELLER, A.H.; TELLER, E. Equation of State Calculations by Fast Computing Machines. Journal of Chemical Physic, 21:1087-1091, 1953.
- [29] HASTINGS, W.K. Monte Carlo Sampling Methods using Markov Chains and their Applications. Biometrika, 55:97-109, 1970.
- [30] MIGON, H.S.; GAMERMAN, D. Statistical Inference: An Integrated Approach. Arnold: New York, 1999.
- [31] BOX, G.E.P; TIAO, G.C. Bayesian Inference in Statistical Analysis. Addison-Wesley, 1972.
- [32] DOUCET, A.; DE FREITAS, N.; GORDON, N. Sequential Monte Carlo Methods in Practice. Springer, 2001.
- [33] TANISAKI H. Nonlinear Filters: Estimation and Applications, Springer, 1996.
- [34] BERGMAN, N. Recursive Bayesian Estimation: Navigation and Tracking Application. PhD Thesis, Department of Electrical Engineering, Linkping University, Sweden, 1999.
- [35] DE FREITAS, N. Bayesian Methods for Neural Networks. PhD Thesis, Trinity College, University of Cambridge, 1999.
- [36] GEWEKE, J. Bayesian Inference in Econometrics Models Using Monte Carlo Integration. Econometrica 24: 1317-1399. 1989

- [37] GORDON, N.J. Bayesian Methods for Tracking. PhD thesis, Imperial College, University of London, 1994
- [38] GORDON, N.J.; SALMOND, D.J.; SMITH, A.F.M. Novel Approach to Nonlinear/NonGaussian Bayesian Estimation. IEEE-Proceedings-F, 140: 107-113, 1993.
- [39] K CRISAN, D.; DOUCET, A. Convergence of Sequential Monte Carlo Methods. Technical Report 381, CUED-F-INFENG, 2000.
- [40] DEL MORAL, P.; LEDOUX, M. Convergence of Empirical Process for Interacting Particle Systems with Application to Non-linear filtering. Journal of Theoretical Probability, 2000.
- [41] TEKALP A.M. Digital Video Processing. Prentice Hall PTR
- [42] JOHNSON R.A., WICHERN D.W. Applied Multivariate Statistical Analysis, Fourth Edition. Prentice Hall. 1998.
- [43] BISHOP C.M. Bayesian PCA, Microsoft Research. St. George House, 1 Guidhall street, Cambridge CB2 3NH, U.K. [cmbichop@microsoft.com]
- [44] BISHOP C.M. Neural Network for Pattern Recognition, Oxford University Press Inc. 1995.
- [45] MACKAY D.J.C. Bayesian Methods for Adaptive Models, Ph.D. thesis, California Institute of Technology. 1992.
- [46] MACKAY D.J.C. Bayesian Non-Linear Modelling with Neural Networks, Cavendish Laboratory, Cambridge, UK, 1995.
- [47] HAYKIN S. Neural Networks A Comprehensive Foundation. Prentice Hall. Second Edition
- [48] WILLIAN, P.M. Bayesian Regularization and pruning using a Laplace prior. Neural Computation 7 (1) 117-143. 1995
- [49] BUNTINE, W.L. and A.S. WEIGEND, Bayesian Back-propagation, Complex System 5, 603-643. 1991.
- [50] NEAL, R.M. Bayesian Learning for Neural Networks, PhD. Thesis, University of Toronto, Canada, 1994.
- [51] TITO, EA.H. Aprendizado Bayesiano Para Redes Neurais, Dissertação de Mestrado, Departamento de Engenharia Elétrica, PUC/Rio, 1999.
- [52] WOLDPERT, D.H. On the use of evidence in neural networks. In S.J. Hanson, J.D. Cowan, and C.L. Giles (Eds.), Advances in Neural Information Processing Systems. Volume 5, pp. 539-546. San Mateo, CA:Morgan Kaufmann. 1993.
- [53] MACKAY, D.J.C. Hyperparameters: optimize or integrate out? In G. Heidbreder (Ed.), Maximun Entropy and Bayesian Methods, Santa Barbara 1993. Dordrech:Kluwer. 1994.
- [54] FAUGERAS O. Three-Dimensional Computer Vision: A Geometric Viewpoint, The MIT Press, Massachusetts, USA, 1993.
- [55] KORN, B.K.P., Robot Vision, The MIT Press, Massachusetts, USA, 1986.

- [56] NABNEY, I.T. NETLAB, Algorithms for Pattern recognition, Springer, 2004.
- [57] THOMAS J.E. Fundamentos de Engenharia de Petróleo. Editora Interciência 2001.

